Elementos de Matemáticas Discretas

C. L. Liu

Segunda Edición

ELEMENTOS DE MATEMÁTICAS DISCRETAS

ELEMENTOS DE MATEMÁTICAS DISCRETAS

Segunda edición

C. L. Liu

Departamento de ciencias de la computación
University of Illinois at Urbana-Champaign

Traducción:
M. en C. Luis Armando Díaz Torres
Instituto Tecnológico de León
Profesor e investigador titular
Ciencias básicas

Revisión técnica:
M. en C. Germán Téllez Castillo
UAM Atzcapotzalco
Departamento de sistemas

McGRAW-HILL

MÉXICO • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA • MADRID • NUEVA YORK
PANAMÁ • SAN JUAN • SANTAFÉ DE BOGOTÁ • SANTIAGO • SÃO PAULO
AUCKLAND • HAMBURGO • LONDRES • MILÁN • MONTREAL • NUEVA DELHI • PARÍS
SAN FRANCISCO • SINGAPUR • ST. LOUIS • SIDNEY • TOKIO • TORONTO

Gerente de producto: Alfonso García Bada M.
Supervisor de edición: Mateo Miguel García
Supervisor de producción: Zeferino García García

ELEMENTOS DE MATEMÁTICAS DISCRETAS

McGRAW-HILL/INTERAMERICANA DE MÉXICO, S. A. de C. V.
Atlacomulco 499-501, Fracc. Ind. San Andrés Atoto,
53500 Naucalpan de Juárez, Edo. de México
Miembro de la Cámara Nacional de la Industria Editorial, Reg. Núm. 1890

ISBN 970-10-0743-3

Traducido de la segunda edición en inglés de
ELEMENTS OF DISCRETE MATHEMATICS

ISBN 0-07-100544-7

1234567890 IP. 95 9087643215

Impreso en México Printed in México

Esta obra se terminó de
imprimir en Junio de 1995 en
Impresora Publi-Mex, S.A. de C.V.
Calz. San Lorenzo 279-32
Delegación Iztapalapa
09850 México, D.F.

Se tiraron 5,000 ejemplares

CONTENIDO

* Todas las secciones marcadas con asterisco pueden omitirse sin afectar la continuidad.

PREFACIO A LA SEGUNDA EDICIÓN

La segunda edición mantiene casi todo el material de la primera edición e incluye tres nuevos capítulos, a saber, el capítulo 2: Computabilidad y lenguajes formales, el capítulo 7: Máquinas de estado finito, y el capítulo 8: análisis de algoritmos, así como también varias secciones nuevas sobre probabilidad discreta, comportamiento asintótico de funciones, y algoritmos recursivos. Espero que el nuevo material sea de utilidad para dar más énfasis a la relevancia de las matemáticas que tratamos de enseñar y que dé al lector una idea acerca de los cursos avanzados en carreras como ciencias de la computación o matemáticas, tales como análisis de algoritmos, teoría de autómatas, lenguajes formales y teoría de probabilidad, temas que el lector desearía llevar una vez que haya comprendido el material de este libro (desde luego, cursos como matemáticas combinatorias, teoría de grafos y álgebra abstracta son continuación natural sin necesidad de material nuevo de esta edición).

Esta edición probablemente contiene más material del que uno puede cubrir en un curso de un semestre a un ritmo promedio. Existen varias posibilidades de adecuar el libro para un curso semestral o un curso cuatrimestral: omitir el análisis sobre sistemas algebraicos saltándose los capítulos 11 y 12, dar un tratamiento más elemental sobre combinatorias evitándose el tópico de funciones generatrices el capítulo 9 y el tópico sobre la solución de relaciones de recurrencia en el capítulo 10, o hacer un breve tratamiento sobre teoría de grafos, limitando la presentación en los capítulos 5 y 6 sólo a definiciones básicas y resultados. En cuanto al orden de presentación, además de seguir el orden actual de los capítulos, se puede posponer el capítulo 3 hasta después del capítulo 8. El material del capítulo 3, aun cuando es elemental, puede aparecer un tanto complicado al estudiante principiante. Los capítulos 3, 9 y 10 hablan sobre tópicos en combinatoria. Por las mismas razones, puede ser conveniente abordar el contenido de los capítulos 2 y 7 juntos, y extenderse en teoría de autómatas y lenguajes formales. Es posible trabajar con el(las) álgebra(s) booleana(s) inmediatamente después de que las proposiciones sean introducidas.

No obstante, el capítulo 12 puede ajustarse si se le cubre inmediatamente después del capítulo 1, dado que éste quizá sería demasiado formal y abstracto para un estudiante (debo además mencionar que el actual ordenamiento de los capítulos está basado más o menos sobre la idea de ir de conjuntos a relaciones, a grafos, a funciones y a estructuras algebraicas). Además de las personas a quienes agradezco en el prefacio de la primera edición quiero también agradecer a R. V. Book, D. H. Haussler, K. S. Khim, H. W. Leong, W. J. Li, M. C. Loui, R. Ness-Cohen, K. H. Pun, W. L. Scherlis, J. Waxman y H. Wu por sus contribuciones.

C. L. Liu

PREFACIO A LA PRIMERA EDICIÓN

Este libro presenta una selección de tópicos como teoría de conjuntos, combinatoria, teoría de grafos y álgebra los cuales considero básicos y útiles para estudiantes de matemáticas aplicadas, ciencias de la computación e ingeniería. Este libro está pensado como texto para un curso en matemáticas discretas a nivel medio universitario, aun cuando también puede ser utilizado en los primeros cursos universitarios, dado que la presentación no supone ningún conocimiento previo más allá de las matemáticas de preparatoria. El material de este libro puede ser cubierto en un curso de un semestre a paso acelerado. Por otra parte, es posible omitir algunos tópicos para un curso algo más holgado. Las secciones marcadas con un asterisco pueden omitirse sin ocasionar problemas en la continuidad.

Este libro es el resultado de un conjunto de apuntes que escribí durante un curso impartido en el Departamento de Ciencias de la Computación de la Universidad de Illinois en Urbana-Champaign. Espero no sea sólo una repetición de lo que abarqué entonces, sino que también sea una reflexión de cómo impartí el material en el salón de clase. He tratado de ser riguroso y preciso en la presentación de los conceptos matemáticos y de evitar formalismos y notaciones complicadas. Como regla, no establecí ninguna definición o hecho si no podía ilustrarlo posteriormente en forma significativa. Así, es muy probable que haya omitido algunas definiciones o hechos "importantes" en este libro. Confío, no obstante, en que los estudiantes serán realmente capaces de encontrar tales definiciones y hechos en algún otro lugar cuando éstos sean requeridos. He intentado enseñar a mis estudiantes algunas matemáticas útiles, de una manera interesante y motivadora. Espero haberles demostrado cómo las matemáticas pueden ser aplicadas para resolver problemas difíciles de la vida real. Además, espero que mis estudiantes no sólo hayan aprendido en el curso algunas herramientas matemáticas muy útiles sino que también desarrollaran sus habilidades para percibir, formular y resolver problemas matemáticos. He tratado de llevar un punto de vista algorítmico en el tratamiento de varios tópicos, si bien decidí no incluir programas de computadora explícitos, principalmente por el tiempo requerido para ello. Espero que algunas de estas preferencias personales y puntos de vista puedan ser compartidos por el profesor que haga uso de este libro.

Me gustaría agradecer a James N. Snyder, mi jefe de departamento, por su estímulo y apoyo, a Murray Edelberg, Jane W. S. Liu y Andrew H. Sherman por su cuidadosa revisión del manuscrito, a Donald K. Friesen por su colaboración en la preparación del manual para el instructor, y a Edward M. Reingold y F. Francés Yao por sus variadas y útiles sugerencias. Varios años atrás tuve la oportunidad de asistir a una mesa redonda sobre el impacto de la computación en las matemáticas, patrocinado por el Comité de Programas Universitarios en Matemáticas de la Asociación Matemática de América. Me he beneficiado ampliamente de esas discusiones sobre la enseñanza de las matemáticas discretas, y estoy muy en deuda con los miembros de aquella mesa redonda. Agradezco también a Glenna Gochenour, Connie Nosbisch, Judy Watkins y June Wingler por su asistencia mecanográfíca y editorial. Finalmente, agradezco a Kathleen D. Liu por su asistencia en la preparación del índice.

Existe cierta similitud entre este libro y el libro *Introducción a las matemáticas combinatorias* que escribí hace algunos años. En algunas partes he seguido muy de cerca la presentación de *Introducción a las matemáticas combinatorias.*

C. L. Liu

CAPÍTULO

UNO

Conjuntos y proposiciones

1.1 INTRODUCCIÓN

Uno de los temas principales de este libro es el estudio de los objetos discretos y de las relaciones que hay entre ellos. El término *objetos discretos* es bastante general; abarca multitud de cosas, como gente, libros, computadoras, transistores, programas de computadora, y demás. Tanto en la vida personal como en el trabajo técnico nos referimos con frecuencia a estos objetos con afirmaciones como "las personas que están en esta habitación estudian el segundo año de su especialidad en ciencias de la computación", o "todos los libros que compré son novelas policíacas escritas por A. B. Charles", o bien "vamos a escoger y a comprar una computadora que sirva para aplicaciones científicas y de negocios, y cuyo precio no exceda de $ 200 000". Nuestra intención es abstraer algunos conceptos básicos relacionados con los diversos tipos de objetos discretos y adoptar una terminología común para manejarlos.

Un indicio de la posibilidad de efectuar dicha abstracción se hace evidente al observar que las tres afirmaciones tienen "algo" en común. En concreto, en la primera afirmación nos referimos a personas que tienen dos atributos: están en ciencias de la computación y son estudiantes de segundo año; en la segunda, a libros que tienen dos atributos: son novelas policíacas y fueron escritos por A. B. Charles; y en la tercera afirmación nos referimos a computadoras que tienen tres atributos: sirven para aplicaciones científicas, para aplicaciones de negocios y cuestan no más de $ 200 000. Dicho de otra forma, considere el grupo de todos los que están en la especialidad de ciencias de la computación y el grupo de todos los estudiantes de la universidad que están en segundo año. Así, en la primera afirmación nos referimos a los estudiantes que pertenecen a ambos grupos. Considere también la colección de todas las novelas policíacas y la colección de todos los libros escritos por A. B. Charles. En la segunda afirmación hacemos referencia a los libros que pertenecen a ambas colecciones. Finalmente, en la tercera afirmación nos referimos a todas las computadoras que pertenecen a las tres categorías siguientes: las que sirven para aplicaciones científicas, las que sirven para aplicaciones de negocios y las que cuestan no más de $ 200 000.

Nuestro ejemplo ilustra las numerosas ocasiones en que tratamos con varias clases de objetos y deseamos referirnos a aquellos que pertenecen a todas las clases. De manera análoga, podríamos percibir inmediatamente los casos en que nos referimos a objetos que pertenecen a una de varias clases, como en la afirmación "quiero entrevistar a todos los estudiantes que hablen alemán o francés", con la que nos referimos a aquellos estudiantes que pertenecen ya sea al grupo de los estudiantes que hablan alemán, o al grupo de estudiantes que hablan francés.

Iniciemos presentando algunos conceptos y términos básicos de la teoría de conjuntos elemental. Un *conjunto* es una colección de objetos *distintos.* Así, el grupo de todos los estudiantes universitarios de segundo año es un conjunto. También son conjuntos el grupo de todos los estudiantes de ciencias de las computación y el grupo de estudiantes de ciencias de la computación que están en segundo año. Usamos la simbología $\{a, b, c\}$ para denotar el conjunto formado por la colección de los objetos *a, b y c.* A los objetos de un conjunto también se les llama *elementos* o *miembros* del conjunto. Por lo común, daremos nombre a los conjuntos. Por ejemplo, escribimos $S = \{a, b, c\}$ para decir que el conjunto llamado S es la colección de los objetos *a, b* y c; en consecuencia, podemos referirnos indistintamente al conjunto S o al conjunto $\{a, b, c\}$. Otro ejemplo:

Estudiantes-de-segundo-año-de-ciencias-de-la-computación = {López, Medrano, Ruiz Yamamoto, Torres}

(El nombre del conjunto {López, Medrano, Ruiz, Yamamoto, Torres}, que es Estudiantes-de-segundo-año-de-ciencias-de-la-computación, es bastante largo. El lector quizá desearía sugerir otros nombres, como S o *SC.* Sin embargo, no hay error conceptual al tener un nombre "largo".) Usamos la notación $a \in$ S para indicar que a es un elemento que está en el conjunto S. En ese caso también decimos que S contiene al elemento *a.* Usamos la notación $d \notin S$ para indicar que d no es un elemento del conjunto S. En ese caso decimos que S no contiene al elemento *d.* Así, en el ejemplo anterior, Medrano $\in$ Estudiantes-de-segundo-año-de-ciencias-de-la-computación, mientras que Rosales $\notin$ Estudiantes-de-segundo-año-de-ciencias-de-la-computación.

Observamos que un conjunto sólo contiene elementos distintos. Así, $\{a, a, b, c\}$ es una representación redundante del conjunto $\{a, b, c\}$. De manera análoga, {El-visitante-de-media-noche, El-visitante-de-media-noche, El-testigo-desaparecido, Main-Street-114} es una representación redundante de las novelas policíacas escritas por A. B. Charles. Podríamos preguntarnos: ¿qué deberemos hacer si tenemos dos ejemplares del libro El-visitante-de-media-noche en nuestra colección de novelas policíacas de A. B. Charles? En ese caso, el conjunto {El-visitante-de-media-noche, El-testigo-desaparecido, Main-Street-114} es el conjunto de títulos distintos de novelas policíacas de A. B. Charles que tenemos en la biblioteca, mientras que el conjunto {El-visitante-de-media-noche-1, El-visitante-de-media-noche-2, El-testigo-desaparecido, Main-Street-114} es el conjunto de novelas policíacas de A. B. Charles que tenemos en la biblioteca, donde El-visitante-de-media-noche-1 es el ejemplar 1 del libro El-visitante-de-media-noche, y El-visitante-de-media-noche-2 es el ejemplar 2 del libro. Dése cuenta que El-visitante-de-media-noche-1 y El-visitante-de-media-noche^ son dos elementos distintos del último conjunto.

También, que los elementos de un conjunto no tienen ningún tipo de orden. Así, *{a, b, c}* y *{b, a, c}* representan la misma colección de elementos. En el capítulo 2 presentaremos el concepto de *conjuntos ordenados.*

Como se mencionó, una manera de decir cuáles son los elementos de un conjunto es listarlos todos. En muchos casos, cuando todos los elementos comparten algunas propiedades, podemos describirlos enunciando las propiedades que caracterizan de manera exclusiva a los que pertenecen a ese conjunto. Por ejemplo, sea $S = \{2,4,6, 8, 10\}$. También podemos especificar los elementos de S diciendo que S es el conjunto de todos los enteros positivos pares no mayores que 10. En efecto, para el conjunto $S = \{2, 4, 6, 8, 10\}$ podemos usar la notación

$$S = \{x \mid x \text{ es un entero positivo par no mayor que } 10\}$$

para el conjunto $\{2,4, 6, 8, 10\}$. En general, usamos la notación

$$\{x \mid x \text{ posee cierta propiedad}\}$$

para un conjunto de objetos que comparten algunas propiedades comunes. Así, S

= {López, Medrano, Ruiz, Yamamoto, Torres}

y

$$S = \{x \mid x \text{ es un estudiante de segundo año de ciencias de la computación}\}$$

son dos maneras diferentes de describir el mismo conjunto de elementos.

Debe señalarse que nuestra definición de conjunto no excluye la posibilidad de tener un conjunto *sin* elementos. El conjunto que no contiene elementos se conoce como *conjunto vacío,* y se denota con { }. (Seguimos respetando la notación de usar un par de llaves para encerrar todos los elementos del conjunto, sólo que en este caso no hay elementos encerrados.) En muchos libros el conjunto vacío también se denota con $\varnothing$. Para que el lector se familiarice con ambas notaciones, usaremos las dos de manera indistinta. Por ejemplo, denotemos con S el conjunto de todas las novelas policíacas de A. B. Charles publicadas en 1924. Resulta claro que S es el conjunto vacío en caso de que A. B. Charles haya nacido en 1925. Otro ejemplo: denotemos con S el conjunto de todos los estudiantes que reprobaron el curso de matemáticas discretas. S podría ser el conjunto vacío si todos los alumnos estudiaron bastante para el examen final.

Advirtamos que no hay restricciones para los elementos de un conjunto. Así, 5= {López, El-visitante-de-media-noche, CDC-6600} es un conjunto bien definido. El hecho de que los elementos López (una persona), El-visitante-de-media-noche (título de un libro) y CDC-6600 (una computadora) parezcan no tener algo en común no impide que sean elementos del mismo conjunto. En efecto, debemos señalar que es correcto tener conjuntos como miembros de un conjunto. Así, por ejemplo, el conjunto *{{a, b, c), d)* contiene los dos elementos *{a, b, c} y d.* El conjunto *{{a, b, c}, a, b, c}* contiene los cuatro elementos *{a, b, c}, a, by c*. El conjunto de todos los comités del senado de Estados Unidos se podría representar con *{{a, b, c}, {a, d,e,f}, {b, e, g}}*, donde cada elemento del conjunto es un comité que, a su vez, es un conjunto cuyos elementos son los senadores que forman ese comité. De manera análoga, *{a, {a}, {{a}}}* es un conjunto con tres elementos *distintos: a*, *{a}* y {{a}}. Además, el conjunto {{ }}, que también se puede expresar como $\{\varnothing\}$, contiene un elemento: el conjunto vacío. El conjunto {{ }, {{ }}}, que también se puede expresar como $\{\varnothing, \{\varnothing\}\}$, contiene dos elementos, el conjunto vacío y un conjunto cuyo único elemento es el conjunto vacío. Quizá sea útil una analogía. Podemos imaginar que *{a, b, c}* corresponde a una "caja" donde hay tres objetos, *a,byc*. Entonces, *{a, b, c, {a,b}}* corresponde a una caja donde hay cuatro objetos: *a,b,c* y una caja, en la cual hay dos objetos, *a* y *b*. También, { } corresponde a una caja vacía; {{ }} corresponde a una caja en la que hay un objeto que resulta ser una caja vacía; y {{ },{{ }}} corresponde a una caja en la que hay dos cajas, una vacía y la otra no. Un ejemplo más: sea

$$S_1 = \{\text{Juan, María}\}$$
$$S_2 = \{\{\text{Juan, María}\}\}$$
$$S_3 = \{\{\{\text{Juan, María}\}\}\}$$

Podemos observar que

$$\begin{aligned} \text{Juan} &\in S_1 \\ \text{Juan} &\notin S_2 \\ \text{Juan} &\notin S_3 \\ S_1 &\in S_2 \\ S_1 &\notin S_3 \\ S_2 &\in S_3 \end{aligned}$$

A partir de dos conjuntos, *PyQ*, decimos que *P* es un *subconjunto* de *Q* si todo elemento de *P* es también un elemento de *Q*. Usaremos la notación $P \subseteq Q$ para indicar que *P* es un subconjunto de *Q*. Por ejemplo, el conjunto *{a, b)* es un subconjunto del conjunto *{y, x, b, c, a}* pero no es subconjunto del conjunto *{a, c, d, e)*. El conjunto de los estudiantes de segundo año de la especialidad de ciencias de la computación es un subconjunto de todos los estudiantes de segundo año. También es un subconjunto de los estudiantes de la especialidad de ciencias de la computación. Por otro lado, el conjunto de los estudiantes de la especialidad de ciencias de la computación no es un subconjunto de todos los estudiantes de segundo aflo, ni el conjunto de todos los estudiantes de segundo año es un subconjunto de todos los estudiantes de la especialidad de ciencias de la computación. Sea *A = {a, b, c}* *y B = {{a, b, c}, a, b, c}*. En efecto, es posible tener $A \in B$ y $A \subseteq B$. Como ejercicio pedimos al lector que verifique las afirmaciones siguientes:

1. Para cualquier conjunto *P*, *P* es un subconjunto de *P*.
2. El conjunto vacío es subconjunto de cualquier conjunto. Sin embargo, el conjunto vacío no necesariamente es elemento de cualquier conjunto.

3. El conjunto $\{\varnothing\}$ no es subconjunto del conjunto $\{\{\varnothing\}\}$, aunque es un elemento del conjunto $\{\{\varnothing\}\}$.

Se dice que dos conjuntos, P y Q, son *iguales* si contienen la misma colección de elementos. Por ejemplo, los dos conjuntos

$$P = \{x \mid x \text{ es un entero positivo par no mayor que } 10\}$$
$$Q = \{x \mid x = y + z \text{ donde } y \in \{1, 3, 5\}, z \in \{1, 3, 5\}\}$$

son iguales. Aunque parezca redundante, podemos decir que dos conjuntos, P y Q, son iguales si P es subconjunto de Q, y Q es subconjunto de P. Más adelante veremos que en ciertas ocasiones ésta es la mejor manera de verificar que dos conjuntos son iguales.

Sea P un subconjunto de Q. Decimos que i[5] es un subconjunto *propio* de Q si P no es igual a Q, esto es, si existe al menos un elemento de Q que no está en P. Por ejemplo, el conjunto $\{a, b\}$ es un subconjunto propio del conjunto $\{y, x, b, c, a\}$. Usamos la notación $P \subset Q$ para indicar que P es un subconjunto propio de Q.

1.2 COMBINACIONES DE CONJUNTOS

Ahora veremos cómo es posible *combinar* conjuntos de varias maneras para obtener nuevos conjuntos. Por ejemplo, sea P el conjunto de estudiantes que toman el curso de teoría de la computación, y sea Q el conjunto de estudiantes que toman el curso de apreciación musical. Si se hizo cierto anuncio, tanto en la clase de teoría de la computación como en la de apreciación musical, ¿cuál es el conjunto de estudiantes que saben del anuncio? Obviamente, es el conjunto de estudiantes que toman ya sea teoría de la computación o apreciación musical, o ambas materias. Si los exámenes finales de ambos cursos están programados para la misma hora, ¿cuál es el conjunto de estudiantes que tendrán dificultades al presentar sus exámenes? Por supuesto, el conjunto de estudiantes que llevan los dos cursos: teoría de la computación y apreciación musical. Para formalizar estos conceptos definimos la unión y la intersección de conjuntos. La *unión* de dos conjuntos, P y Q, denotada por $P \cup Q$, es el conjunto de aquellos elementos que son exactamente los elementos de P o de Q, *o* de ambos.[1'] Por ejemplo,

$$\{a, b\} \cup \{c, d\} = \{a, b, c, d\}$$
$$\{a, b\} \cup \{a, c\} = \{a, b, c\}$$
$$\{a, b\} \cup \varnothing = \{a, b\}$$
$$\{a, b\} \cup \{\{a, b\}\} = \{a, b, \{a, b\}\}$$

† Esperaremos hasta el capítulo 11 para definir el concepto de operaciones algebraicas. Por el momento, $P \cup Q$ es, simplemente, el nombre que hemos escogido para un conjunto.

La *intersección* de dos conjuntos, P y Q, denotada por $P \cap Q$, es el conjunto cuyos elementos son exactamente los elementos que están tanto en P como en Q. Por ejemplo,

$$\begin{aligned} \{a, b\} \cap \{a, c\} &= \{a\} \\ \{a, b\} \cap \{c, d\} &= \varnothing^{\dagger} \\ \{a, b\} \cap \varnothing &= \varnothing \end{aligned}$$

Si los elementos de P están caracterizados por una propiedad común y los elementos de Q por otra propiedad común; entonces, la unión de P y Q es el conjunto de elementos que poseen al menos una de estas propiedades, y la intersección de P y Q es el conjunto de elementos que poseen ambas propiedades. De acuerdo con las definiciones, $P \cup Q$ y $Q \cup P$ denotan al mismo conjunto, y lo mismo sucede con $P \cap Q$ y $Q \cap P$.

Por tanto, la unión del conjunto $P \cup Q$ y el conjunto R, denotada por $(P \cup Q) \cup R$, donde los paréntesis se usan como delimitadores para evitar confusión, contiene exactamente los elementos de P, los elementos de Q y los elementos de R. Usaremos la notación $P \cup Q \cup R$ en lugar de $(P \cup Q) \cup R$, y nos referiremos al conjunto $P \cup Q \cup R$ como a la unión de los tres conjuntos P, Q y R. En general, la unión del conjunto $(\ldots((P_1 \cup P_2) \cup P_3)\ldots) \cup P_{k-1}$ y el conjunto P_k, denotada por $(\ldots((P_1 \cup P_2) \cup P_3)\ldots \cup P_{k-1}) \cup P_k$, contiene exactamente los elementos de P_1, los elementos de $P_2, \ldots$, los elementos de P_{k-1}, y los elementos de P_k. Usaremos la notación $P_1 \cup P_2 \cup P_3 \cdots \cup P_{k-1} \cup P_k$ en lugar de $(\ldots((P_1 \cup P_2) \cup P_3)\ldots \cup P_{k-1} \cup P_k$ y nos referiremos al conjunto $P_1 \cup P_2 \cup P_3 \ \ldots \cup P_{k-1} \cup P_k$ como a la unión de los k conjuntos $P_1, P_2, P_3, \ldots, P_{k-1}, P_k$. De manera análoga, la intersección del conjunto $P \cap Q$ y el conjunto R, denotada por $(P \cap Q) \cap R$, contiene exactamente los elementos que están en P, en Q y en R. Además, la intersección del conjunto $(\ldots((P_1 \cap P_2) \cap P_3)\ldots) \cap P_{k-1}$ y el conjunto P_k, denotada por $(\ldots((P_1 \cap P_2) \cap P_3)\ldots \cap P_{k-1} \cap P_k$, contiene exactamente los elementos que están en P_1, en $P_2, \ldots$, en P_{k-1}, y en P_k. Usaremos la notación $P_1 \cap P_2 \cap P_3 \cdots \cap P_{k-1} \cap P_k$ en lugar de $(\ldots((P_1 \cap P_2) \cap P_3)\ldots \cap P_{k-1}) \cap P_k$ y nos referiremos al conjunto $P_1 \cap P_2 \cap P_3 \cdots \cap P_{k-1} \cap P_k$ como a la intersección de los k conjuntos $P_1, P_2, P_3, \ldots, P_{k-1}, P_k$. Pongamos el siguiente ejemplo: el conjunto de los estudiantes de licenciatura de una universidad es la unión de los conjuntos de los estudiantes de primer año, los de segundo, los de tercero y los de cuarto, mientras que el conjunto de los estudiantes que van a graduarse es la intersección del conjunto de los de cuarto año, el conjunto de los estudiantes que han acumulado 144 horas crédito, o más, y el de los estudiantes que tienen promedio de C, o mejor.

Denotemos con P el conjunto de los estudiantes que toman teoría de la computación, con Q el conjunto de estudiantes que toman apreciación musical, y con R el conjunto de estudiantes que tienen tipo de sangre AB. Suponga que en las clases de teoría de la computación y de apreciación musical se hace un llamado de emergencia solicitando donadores de sangre tipo AB. Queremos determinar cuáles son los miembros del conjunto de donadores potenciales que escucharon el llamado de emergencia. Como $S = P \cup Q$ es el conjunto de estudiantes que escucharon el llamado de emergencia, $R \cap S$ será el conjunto de donadores potenciales que escucharon el llamado de emergencia. En lugar de usar el

† Se dice que dos conjuntos son *ajenos* cuando su intersección es el conjunto vacío.

nuevo nombre S para el conjunto $P \cup Q$, podemos escribir simplemente $R \cap (P \cup Q)$. Observe que el conjunto de donadores potenciales que escucharon la llamada de emergencia también es el conjunto de estudiantes con sangre de tipo AB que están en la clase de teoría de la computación, y el conjunto de estudiantes con sangre de tipo AB que están en la clase de apreciación musical, esto es, el conjunto $(R \cap P) \cup (R \cap Q)$. Este ejemplo sugiere que para cualesquiera conjuntos P, Q y R, los dos conjuntos $R \cap (P \cup Q)$ y $(R \cap P) \cup (R \cap Q)$ son iguales. Así sucede, en efecto, como lo vamos a demostrar.

Primero demostraremos que $R \cap (P \cup Q)$ es un subconjunto de $(R \cap P) \cup (R \cap Q)$ al ver que todo elemento de $R \cap (P \cup Q)$ está también en $(R \cap P) \cup (R \cap Q)$. Sea x un elemento de $R \cap (P \cup Q)$. El elemento x debe estar en R y debe estar en P o en Q. Si x está en P, x está en $R \cap P$. Si x está en Q, x está en $R \cap Q$. En consecuencia, x está en $(R \cap P) \cup (R \cap Q)$, y concluimos que $R \cap (P \cup Q)$ es un subconjunto de $(R \cap P) \cup (R \cap Q)$. En segundo lugar demostraremos que $(R \cap P) \cup (R \cap Q)$ es un subconjunto de $R \cap (P \cup Q)$. Sea x un elemento de $(R \cap P) \cup (R \cap Q)$. Así, x debe estar ya sea en $R \cap P$ o en $R \cap Q$. Esto es, x debe estar ya sea en R y P o en R y Q. En otras palabras, x debe estar en R y debe estar ya sea en P o en Q. En consecuencia, x está en $R \cap (P \cup Q)$, y podemos concluir que $(R \cap P) \cup (R \cap Q)$ es un subconjunto de $R \cap (P \cup Q)$. Se sigue que los conjuntos $R \cap (P \cup Q)$ y $(R \cap P) \cup (R \cap Q)$ son iguales.

De manera análoga podemos demostrar que para cualesquiera conjuntos P, Q y R, los dos conjuntos $R \cup (P \cap Q)$ y $(R \cup P) \cap (R \cup Q)$ son iguales. Más aún, tenemos que

$$R \cap (P_1 \cup P_2 \cup \cdots \cup P_k) = (R \cap P_1) \cup (R \cap P_2) \cup \cdots \cup (R \cap P_k)$$
$$R \cup (P_1 \cap P_2 \cap \cdots \cap P_k) = (R \cup P_1) \cap (R \cup P_2) \cap \cdots \cap (R \cup P_k)$$

Dejamos los detalles al lector.†

La *diferencia* de dos conjuntos, P y Q, denotada con $P - Q$, es el conjunto que contiene exactamente aquellos elementos de P que no están en Q. Por ejemplo,

$$\{a, b, c\} - \{a\} = \{b, c\}$$
$$\{a, b, c\} - \{a, d\} = \{b, c\}$$
$$\{a, b, c\} - \{d, e\} = \{a, b, c\}$$

Si P es el conjunto de personas que tienen boletos para un juego de pelota y Q es el conjunto de personas que están enfermas el día del juego, entonces $P - Q$ es el conjunto de personas que irán al juego. Note que Q puede contener alguno o ninguno de los elementos del conjunto P. Sin embargo, estos elementos de ningún modo aparecen en $P - Q$, así como en el ejemplo, aquellas personas que están enfermas pero que no tienen boletos para el juego de pelota, de todos modos no irán al juego de pelota. En efecto, si los elementos de Q están caracterizados por alguna propiedad común, entonces $P - Q$ es el conjunto de elementos de P que no poseen

† Insistimos en no introducir los conceptos de operaciones algebraicas, asociatividad y distributividad hasta el capítulo 11. Observe, sin embargo, que estos conceptos no se necesitan aquí, pues $P \cap Q$, $P_1 \cap P_2 \cap \cdots \cap P_k$, $P \cup Q$ y $P_1 \cup P_2 \cup \cdots \cup P_k$ son, simplemente, nombres de conjuntos obtenidos a partir de nuestras definiciones.

esta propiedad. Si Q es un subconjunto de P, el conjunto $P - Q$ se llama también *complemento de Q respecto a P*. Por ejemplo, sea P el conjunto de todos los estudiantes del curso de teoría de la computación y Q el de los estudiantes que han aprobado el curso. Entonces $P - Q$ es el conjunto de los estudiantes que lo han reprobado. En ocasiones, cuando es claro a qué contexto pertenece el conjunto P, abreviaremos el *complemento de Q respecto a P* como *el complemento de Q*, que denotaremos con $\bar{Q}$. Por ejemplo, sea P el conjunto de todos los estudiantes del curso de teoría de la computación; Q el conjunto de los inscritos en la especialidad de ciencias de la computación que toman el curso, y R el conjunto de los estudiantes de segundo año que toman el curso. Entonces, el complemento de Q se refiere al conjunto de los estudiantes del curso que no están en la especialidad de ciencias de la computación, y el complemento de R al conjunto de los estudiantes que no están en segundo año, si se sobreentiende que restringimos nuestro análisis a los estudiantes que están en el curso de teoría de la computación. En efecto, siempre que nuestro análisis se restringe a los subconjuntos de un conjunto P, nos referimos a P como al *universo*.

La *diferencia simétrica* de dos conjuntos, P y Q, denotada por $P \oplus Q$, es el conjunto que contiene exactamente todos los elementos que están en P o en Q, pero no en ambos. En otras palabras, $P \oplus Q$ es el conjunto $(P \cup Q) - (P \cap Q)$. Por ejemplo,

$$\{a, b\} \oplus \{a, c\} = \{b, c\}$$
$$\{a,b\} \oplus \varnothing = \{a,b\}$$
$$\{a, b\} \oplus \{a, b\} = \varnothing$$

Si denotamos con P al conjunto de autos que tienen la dirección defectuosa y con Q al conjunto de autos que tienen la transmisión defectuosa, entonces $P \oplus Q$ es el conjunto de autos que tienen uno de los defectos, pero no ambos. Supongamos que un estudiante obtendrá una A en un curso si tuvo un buen desempeño en los dos exámenes parciales, tendrá B si salió bien en alguno de los dos exámenes, y tendrá C si tuvo un bajo desempeño en ambos exámenes. Sea P el conjunto de los estudiantes que tuvieron buen desempeño en el primer examen y Q el conjunto de estudiantes que tuvieron buen desempeño en el segundo examen. Entonces $P \cap Q$ es el conjunto de estudiantes que tendrán A, $P \oplus Q$ es el conjunto de estudiantes que obtendrán B, y $S - (P \cup Q)$ es el conjunto de estudiantes que tendrán C, donde S es el conjunto de todos los estudiantes del curso. Definimos $P_1 \oplus P_2 \oplus \cdots \oplus P_k$ como el conjunto de elementos que están en un número impar de los conjuntos $P_1, P_2, \ldots, P_k$.

El *conjunto potencia* de un conjunto A, denotado con $\mathscr{P}(A)$, es el conjunto que contiene exactamente todos los subconjuntos de A. Así, $\mathscr{P}(\{a, b\}) = \{\{\ \}, \{a\}, \{b\}, \{a, b\}\}$, y $\mathscr{P}(\{\ \}) = \{\{\ \}\}$. Observe que para cualquier conjunto A, $\{\ \} \in \mathscr{P}(A)$, así como $\{\ \} \subseteq \mathscr{P}(A)$. Por ejemplo, sea A = {novela, publicado-en-1975, edición-rústica} el conjunto de tres atributos sobre los libros que nos interesan de la biblioteca. Entonces $\mathscr{P}(A)$ es el conjunto de todas las combinaciones posibles de estos atributos que pueden poseer los libros, que van desde los libros que no tienen estos atributos [el conjunto vacío en $\mathscr{P}(A)$] hasta libros que tienen los tres atributos [el conjunto A en $\mathscr{P}(A)$].

Los conjuntos obtenidos a partir de combinaciones de conjuntos dados se pueden representar de manera gráfica. Si P y Q son los conjuntos representados por las áreas sombreadas de la figura 1.1a, entonces las áreas sombreadas de la figura 1.16 representan los conjuntos $P \cup Q$, $P \cap Q$, $P - Q$ y $P \oplus Q$, respectivamente. Estos diagramas se conocen como *diagramas de Venn*.

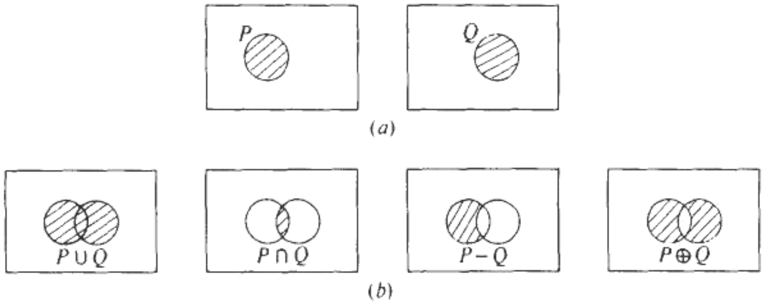


Figura 1.1

1.3 CONJUNTOS FINITOS E INFINITOS

Es bastante claro, intuitivamente, que el tamaño de un conjunto es el número de elementos distintos del conjunto. Así, no habrá duda cuando digamos que el tamaño del conjunto $\{a, b, c)$ es 3, el tamaño del conjunto $\{a, \varnothing, d\}$ también es 3, el tamaño del conjunto $\{\{a, b)\}$ es 1, y el tamaño del conjunto $\varnothing$ es 0. En efecto, este análisis sobre el tamaño de los conjuntos sería suficiente si sólo estuviéramos interesados en el tamaño de conjuntos "finitos". Sin embargo, un tema más misterioso es el del tamaño de los conjuntos "infinitos". A estas alturas, un lector perceptivo quizás haría la pregunta: "En primer lugar, ¿qué es un conjunto infinito?" Una respuesta evasiva como "un conjunto infinito es un conjunto que no es finito", no constituye una respuesta, pues si lo pensamos un poco, también deberíamos responder la pregunta: "Bueno, ¿entonces qué es un conjunto finito?"

Comencemos por establecer que aún no nos proponemos exponer las definiciones precisas de conjuntos finitos y de conjuntos infinitos. Como base de nuestro análisis queremos construir un ejemplo de conjunto infinito. Para un conjunto dado A, definimos el *sucesor* de A, denotado con A^+, como el conjunto $A \cup \{A\}$. Observemos que $\{A\}$ es un conjunto que contiene a A como único elemento. En otras palabras, A^+ es un conjunto que consta de todos los elementos de A junto con un elemento adicional, que es el conjunto A. Por ejemplo, si $A = \{a, b\}$, entonces $A^+ = \{a, b\} \cup \{\{a, b\}\} = \{a, b, \{a, b\}\}$; y si $A = \{\{a\}, b)$, entonces $A^+ = \{\{a\}, b, \{\{a\}, b\}\}$. Ahora construyamos una sucesión de conjuntos a partir del conjunto vacío $\varnothing$. El sucesor del conjunto vacío es $\{\varnothing\}$, cuyo sucesor es $\{\varnothing, \{\varnothing\}\}$, y cuyo sucesor es, a su vez, $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$. Es claro que podemos continuar y construir más y más sucesores. Asignemos nombres a estos conjuntos. En particular, usaremos 0, 1, 2, 3,... como nombres de los conjuntos.† Sean

$$0 = \varnothing$$
$$1 = \{\varnothing\}$$
$$2 = \{\varnothing, \{\varnothing\}\}$$
$$3 = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$$
$$\cdots\cdots\cdots\cdots\cdots$$

† Es tan correcto usar 0,1,2, 3,... como nombres de conjuntos, como usar A, B, C, D,... Como se verá más adelante, escogemos los nombres 0, 1,2,3,... con toda intención.

Tenemos $1 = 0^+$, $2 = 1^+$, $3 = 2^+$, y así sucesivamente. Ahora definamos un conjunto *N* tal que

1. *N* contiene al conjunto 0.
2. Si el conjunto *n* es un elemento de *N*, entonces también lo es el conjunto $n+$.
3. *N* no contiene otros conjuntos.

Como para todo conjunto en *N* se tiene que su sucesor también está en *N*, es probable que el lector admita que, en efecto, *N* es un "conjunto infinito". Sin embargo, procedamos de una manera más precisa.

Hablaremos de tamaños de conjuntos de manera comparativa. Para ello, introduzcamos una definición: Dados dos conjuntos, *P y Q*, decimos que existe una *correspondencia biunívoca* o *correspondencia uno a uno,* entre los elementos de *P* y los elementos de *Q* si es posible aparear los elementos de *P* y de *Q* de modo tal que todos los elementos de *P* estén apareados con distintos elementos de *Q*† Así, existe una correspondencia uno a uno entre los elementos del conjunto *{a, b)* y los elementos del conjunto *{c, d),* pues podemos aparear *a* con *c* y *b* con *d,* o podemos aparear *a* con *dy b* con *c.* También existe una correspondencia uno a uno entre los elementos del conjunto $\{a, b, c\}$ *y* los elementos del conjunto $\{\varnothing, a, d\}$. Por otro lado, no existe correspondencia biunívoca entre los conjuntos *{a, b, c)* y *{a, d}.* Es clara la intención al introducir el concepto de correspondencia biunívoca, o uno a uno, entre elementos de dos conjuntos, pues ahora podemos comparar dos conjuntos y decir si son del mismo o de diferente tamaño. En efecto, la base para comparar está formada por los conjuntos que construimos anteriormente, a saber, 0, 1, 2, 3,..., y *N.* Ahora estamos preparados para introducir algunas definiciones formales. Se dice que un conjunto es *finito* si existe una correspondencia uno a uno entre los elementos del conjunto y los elementos de algún conjunto *n,* donde $n \in N$; se dice que *n* es la *cardinalidad* del conjunto. Así, por ejemplo, la cardinalidad de los conjuntos $\{a, b, c\}$, $\{a, \varnothing, d\}$ y $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$ es, en todos los casos, igual a 3. Observe que, ahora, decir que un conjunto es infinito si no es un conjunto finito, es algo preciso. Podemos, sin embargo, ser más precisos acerca del "tamaño" de los conjuntos infinitos: se dice que un conjunto es *infinito contable,* o *numerable* (o que la cardinalidad del conjunto es infinita contable (o contablemente infinita))‡ si existe una correspondencia uno a uno entre los elementos del conjunto y los elementos de *N.* Observamos, en primer lugar, que el conjunto de todos los números naturales {0,1,2, 3,.. .}§ es un conjunto infinito contable. Se sigue que el conjunto de todos los enteros pares no negativos {0, 2,4, 6, 8,...} es un conjunto infinito contable, pues existe una correspondencia biunívoca obvia entre todos los enteros pares no negativos y todos los números naturales, a saber, al entero par $2i$ le corresponde el número natural *i*, para $i = 0, 1, 2, \ldots$ De manera análoga, el conjunto de todos los múltiplos de 7 no negativos {0,7,14,21,...} también es un conjunto infinito contable. También lo es el conjunto de todos los enteros positivos {1, 2, 3, ...}. Señalemos que un conjunto es infinito contable si, comenzando con cierto elemento, podemos listar sucesivamente, uno detrás de otro, todos los elementos del conjunto, pues

† Esta definición intuitiva se hará más formal en el capítulo 4.
‡ Los libros sobre el tema también se refieren a la cardinalidad de un conjunto infinito contable como $\aleph_0$ ($\aleph$ es la primera letra en el alfabeto hebreo).
§ Quizá la notación sea algo confusa, pero es intencional porque el conjunto N es, en *efecto,* una definición precisa del conjunto de los números naturales.

esa lista nos permite construir una correspondencia uno a uno entre los elementos del conjunto y los números naturales. Por ejemplo, el conjunto de todos los enteros {..., -2,-1, 0, 1, 2, . . .} es un conjunto infinito contable, porque sus elementos se pueden listar sucesivamente como {0, 1, -1,2, -2, 3,-3,. ..}. Este ejemplo sugiere que la unión de dos conjuntos infinitos contables también es un conjunto infinito contable. En efecto, así es. De hecho, la unión de un número finito de conjuntos infinitos contables es un conjunto infinito contable y, lo que es más, lo mismo sucede con la unión de un número infinito contable de conjuntos infinitos contables (véase el problema 1.26).

1.4 CONJUNTOS INFINITOS NO CONTABLES

En esta sección mostraremos que existen conjuntos infinitos cuya cardinalidad no es infinita contable. Ahora introducimos una "técnica de demostración" que quizá sea nueva para el lector.*

Se pidió a tres niños (Jaime, José y Juan) que probaran helados de tres sabores diferentes, chocolate, vainilla y fresa. La tabla siguiente indica si les gustó o no cada sabor:

	Chocolate	Vainilla	Fresa
Jaime	Sí	No	Sí
José	No	No	Sí
Juan	Sí	Sí	Sí

Hagamos la siguiente observación trivial: supongamos que nos dicen que hay un niño que está en desacuerdo con Jaime acerca de si el helado de chocolate es delicioso (esto es, a Jaime le gusta el helado de chocolate pero al niño no le gusta), está en desacuerdo con José acerca de si el helado de vainilla es delicioso (esto es, a José no le gusta el helado de vainilla pero al niño sí le gusta), y también está en desacuerdo con Juan acerca de si el helado de fresa es delicioso (esto es, a Juan le gusta el helado de fresa pero al niño no le gusta). Como se ve, este niño no puede ser Jaime, ni José, ni Juan, debe ser *otro* niño, ya que está en desacuerdo con los gustos de cada uno en al menos uno de los sabores, según se ilustra en la tabla siguiente:

	Chocolate	Vainilla	Fresa
Jaime	Sí	No	Sí
José	No	No	Sí
Juan	Sí	Sí	Sí
Niño nuevo	No	Sí	No

† En efecto, en este libro necesitaremos demostrar en varias ocasiones que es imposible que algo exista, o que es imposible realizar alguna tarea. Todos sabemos cómo demostrar la existencia de algo, o cómo describir un procedimiento para realizar cierta tarea. Pero, ¿cómo demostrar que es imposible, en cualquier circunstancia, que algo exista? ¿Cómo probar que es imposible realizar una cierta tarea, no importa lo inteligente que sea la persona o lo poderosa que sea la máquina?

Observemos una situación más general. Supongamos que hay *n* niños y helados de *n* sabores. Si nos dijeran que existe un niño que está en desacuerdo con el primer niño acerca de si el primer sabor es delicioso, que está en desacuerdo con el segundo niño acerca de si el segundo sabor es delicioso, y que está en desacuerdo con el *n*-ésimo niño acerca de si el *n*-ésimo sabor es delicioso, entonces podemos estar seguros de que este niño no es uno de los *n* niños, porque está en desacuerdo con cada uno de ellos en al menos un aspecto. Este ejemplo en apariencia frívolo ilustra un *método diagonal* mediante el cual aseguramos que cierto objeto (un niño nuevo) no es uno de los objetos dados (los *n* niños que conocemos), usando el hecho de que este objeto es diferente de cada uno de los objetos dados en al menos un aspecto.

Como ejemplo de los conjuntos infinitos con cardinalidad que no es infinita contable, ahora mostraremos que el conjunto de los números reales entre 0 y 1 no es un conjunto infinito contable. Nuestro método de demostración es suponer que el conjunto es infinito contable y después mostrar que existe una contradicción. Si la cardinalidad del conjunto de los números reales entre 0 y 1 es infinita contable, entonces existe una correspondencia uno a uno entre estos números reales y los números naturales. En consecuencia, podemos listarlos exhaustivamente uno detrás de otro en forma decimal, de la manera siguiente:†

$$
\begin{array}{l}
0.a_{11}a_{12}a_{13}a_{14}\cdots \\
0.a_{21}a_{22}a_{23}a_{24}\cdots \\
0.a_{31}a_{32}a_{33}a_{34}\cdots \\
\cdots\cdots\cdots\cdots\cdots \\
0.a_{i1}a_{i2}a_{i3}a_{i4}\cdots \\
\cdots\cdots\cdots\cdots\cdots
\end{array}
$$

donde a_{ij} denota el j-ésimo dígito del i-ésimo número de la lista. Considere el número

$$0.b_1b_2b_3b_4\cdots$$

donde

$$b_i = \begin{cases} 1 & \text{si} \quad a_{ii} = 9 \\ 9 - a_{ii} & \text{si} \quad a_{ii} = 0, 1, 2, \ldots, 8 \end{cases}$$

para todo *i*. Como se ve, el número $0.b_1b_2b_3b_4\ldots$ es un número real entre 0 y 1 que no tiene una cola infinita de ceros (por ejemplo 0.34000…). Más aún, es distinto de cada uno de los números de la lista anterior porque difiere del primer número en el primer dígito, del segundo número en el segundo dígito, del *i*-ésimo número en el *i*-ésimo dígito, y así sucesivamente. En consecuencia, concluimos que la lista anterior no es un listado exhaustivo del conjunto de todos los números reales entre 0 y 1, contradiciendo la hipótesis de que este conjunto es infinito contable.

Es posible continuar en esta dirección y clasificar los conjuntos infinitos de manera que se puedan precisar conceptos como que algunos conjuntos son "más infinitos" que otros conjuntos infinitos. Esto, sin embargo, rebasa el ámbito de nuestro análisis.

† Un número como 0.34 se puede escribir de dos maneras diferentes, a saber, 0.34000… o 0.339999 …. Seguimos la convención arbitraria de escribirlo de la segunda manera.

1.5 INDUCCIÓN MATEMÁTICA

Primero consideremos algunos ejemplos ilustrativos:

Ejemplo 1.1

Demos por hecho que tenemos timbres postales de dos denominaciones diferentes, 3 centavos y 5 centavos. Queremos demostrar que es posible realizar cualquier envío de 8 centavos, o más, usando únicamente timbres de estas dos denominaciones. Desde luego, el enfoque de mostrar caso por caso cómo formar el envío de 8 centavos, de 9 centavos, de 10 centavos, y así sucesivamente, usando timbres de 3 centavos y de 5 centavos, no será práctico, pues hay un número infinito de casos por examinaré Consideremos un enfoque alternativo. Queremos demostrar que de ser posible hacer un envío de k centavos usando timbres de 3 centavos y de 5 centavos, entonces también es posible hacer un envío de $k + 1$ centavos usando timbres de 3 centavos y de 5 centavos. Examinemos dos casos: supongamos que hacemos un envío de k centavos usando al menos un timbre de 5 centavos. Al remplazar un timbre de 5 centavos por dos timbres de 3 centavos habremos formado un envío de $k+ 1$ centavos. Por otro lado, supongamos que hacemos un envío de k centavos usando solamente timbres de 3 centavos. Como $k \geqq 8$, debe haber al menos tres timbres de 3 centavos. Al remplazar tres timbres de 3 centavos por dos timbres de 5 centavos habremos formado un envío de $k + 1$ centavos. Como la manera de hacer un envío de 8 centavos es obvia, podemos concluir que es posible hacer un envío de 9 centavos, lo que, a su vez, nos conduce a concluir que podemos hacer un envío de 10 centavos, lo que, a su vez, nos conduce a concluir que podemos hacer un envío de 11 centavos, y así sucesivamente. □

Ejemplo 1.2

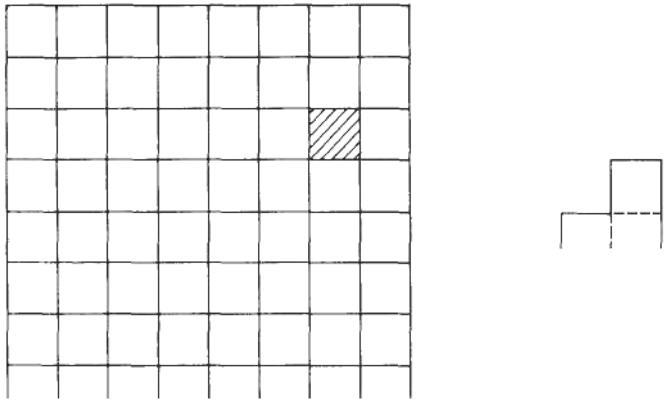
Supongamos que quitamos un cuadrado a un tablero común de ajedrez, de 8 x 8, según se muestra en la figura 1.2*a*. Dados los 21 triminós‡ en forma de L mostrados en la figura 1.2*b*, queremos saber si es posible *cubrir* los 63 cuadrados restantes del tablero con los triminós. (Por *cubrir* los cuadrados restantes del tablero queremos decir que cada cuadrado deberá estar tapado por exactamente uno, sin que salgan del tablero partes de los triminós, ni cubran el cuadro removido.) La respuesta es afirmativa, como lo muestra la figura 1.3. En realidad podemos probar un resultado más general, lo cual haremos a continuación.

Al tablero de ajedrez con uno de sus cuadrados removidos, le llamaremos tablero de ajedrez *defectuoso.* Queremos demostrar que cualquier tablero de ajedrez defectuoso de 2" x 2" se puede cubrir con triminós en forma de L.§ Es obvio que un tablero de ajedrez defectuoso de 2 x 2 se puede cubrir con un triminó en forma de L. Ahora supongamos que cualquier tablero de ajedrez defectuoso de 2^k x 2^k se puede cubrir con triminós en forma de L, y procedamos a demostrar que cualquier tablero de ajedrez defectuoso de $2k^{+1}$ x $2k^{+1}$ también se puede cubrir con triminós en forma de L. Considere

† Véase, sin embargo, el problema 1.30.

‡ La palabra triminó se deriva de la palabra dominó. También hay *tetraminós, pentaminós, hexaminós,* y, en general, *poliminós.* Para ver varios resultados interesantes sobre poliminós consúltese Golomb [7].

§ De inmediato uno se pregunta si 2" x 2" - 1 siempre es divisible entre 3. La respuesta es afirmativa. (Véase el problema 1.36.)



a) *b)*

Figura 1.2

un tablero de ajedrez defectuoso de 2^{k+1} x 2^{k+1} según se muestra en la figura 1.4a. Dividamos el tablero en cuatro cuadrantes, cada uno de los cuales es un tablero de 2^k x 2^k, según se muestra en la figura 1.4*b*. Uno de estos tableros de 2^k x 2^k es defectuoso. Más aún, al colocar un triminó en forma de L en el centro del tablero de 2^{k+1} x 2^{k+1}, como se muestra en la figura 1.4*c*, podemos imaginar que los otros tres cuadrantes también son tableros de ajedrez defectuosos de 2^k x 2^k. Como supusimos que cualquier tablero defectuoso de 2^k x 2^k se puede cubrir con triminós en forma de L, podemos cubrir cada uno de los cuadrantes con triminós en forma de L, y concluir que cualquier tablero de ajedrez defectuoso de 2^{k+1} x 2^{k+1} se puede cubrir con triminós en forma de L. Así, comenzando por cubrir cualquier tablero de ajedrez defectuoso de 2 x 2, hemos probado que podemos cubrir cualquier tablero de ajedrez defectuoso de 2" x 2".

Estos dos ejemplos ilustran una poderosa técnica de demostración en matemáticas, conocida como el principio de *inducción matemática.* Si para una afirmación acerca de un número natural *n* podemos demostrar que:

1. La afirmación es verdadera para $n = n_0$ *; y*
2. La afirmación es verdadera para $n = k + 1$, suponiendo que la afirmación es verdadera para

entonces podemos concluir que la afirmación es verdadera para todos los números naturales

Al punto 1 se le conoce como *base de la inducción* y al 2 como *paso de inducción.* A la suposición en 2 de que la afirmación es verdadera para $n = k$ también se le conoce como *hipótesis de inducción.* Por ejemplo, en el problema de los timbres postales queremos demostrar la afirmación: "Es posible realizar cualquier envío de *n* centavos usando timbres de 3 y 5 centavos, para $n \geqq 8$." Para demostrar la afirmación demostramos que:

1. *Base de la inducción.* Es posible hacer un envío de exactamente 8 centavos.
2. *Paso de inducción.* Es posible hacer un envío de $k + 1$ centavos, suponiendo que es posible formar un envío de exactamente *k* centavos $(k \geqq 8)$.

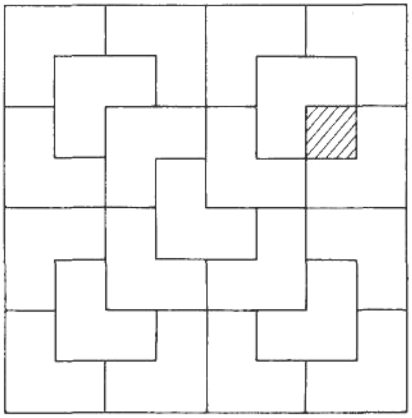
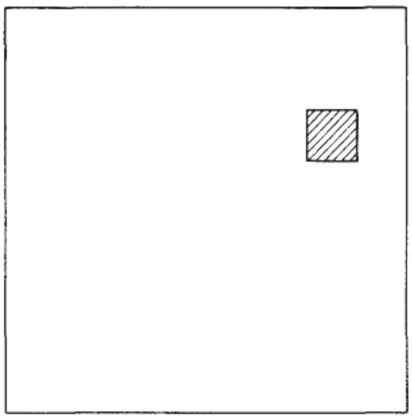
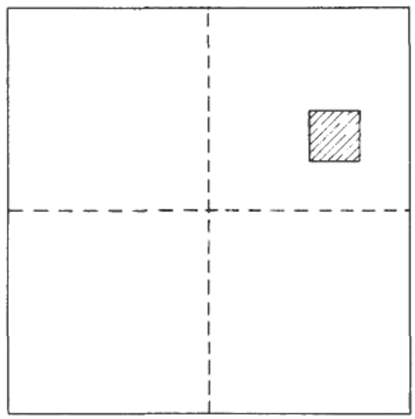


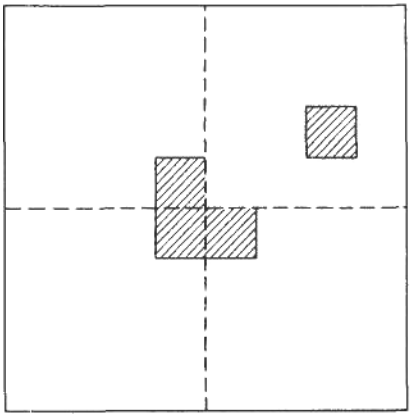
Figura 1.3



a)



b)



c)

Figura 1.4

El principio de inducción matemática es una consecuencia directa de la definición de números naturales. Consideremos un conjunto S tal que:

1. El número natural n_0 está en S.
2. Si el número natural k está en S, entonces el número natural $k + 1$ también está en S, $(k \geqq n_0)$.

De acuerdo con la definición del conjunto de números naturales podemos concluir que S contiene a todos los números naturales mayores o iguales que n_0. Sin embargo, ésta es precisamente la afirmación del principio de inducción matemática, cuando consideramos que S es el conjunto de los números naturales para los que es verdadera una afirmación dada. Veamos más ejemplos:

Ejemplo 1.3

El rey convocó a las mejores matemáticas del reino a presentarse en palacio para ver qué tan listas eran. Les dijo: "A algunas de ustedes les he puesto sombreros blancos y a otras sombreros negros. Pueden verse entre sí pero no pueden hablar. Yo me retiro y vendré cada hora. Cada vez que regrese quiero que aquellas de ustedes que hayan concluido que usan sombrero blanco vengan y me lo digan de inmediato." Conforme pasó el tiempo, a la n-ésima hora cada una de las n matemáticas que usaban sombrero blanco informaron al rey que sabían que usaban sombrero blanco. ¿Por qué?

Probaremos por inducción que si hay n matemáticas que llevan sombrero blanco, entonces todas lo sabrán en la n-ésima hora.

1. *Base de la inducción.* Para $n = 1$, sólo hay una matemática que tiene sombrero blanco. Como el rey dijo que había puesto sombrero blanco en alguna (los reyes nunca mienten), la matemática que vio que todas las demás llevaban sombrero negro, comprendió de inmediato que ella llevaba el sombrero blanco. En conse cuencia, ella informaría al rey en la primera hora (cuando el rey volviera por primera vez) que llevaba sombrero blanco.
2. *Paso de inducción.*† Supongamos que si hubiera k matemáticas con sombrero blanco entonces hubieran comprendido que llevaban sombrero blanco e informado al rey en la k-ésima hora. Ahora supongamos que había $k + 1$ matemáticas con sombrero blanco. Cada matemática con sombrero blanco veía que k de sus colegas llevaban sombrero blanco. Sin embargo, que sus colegas no informaran al rey de su hallazgo en la hora k-ésima sólo puede implicar que había más de k personas

†Para entender mejor el razonamiento, exploremos el caso de dos matemáticas con sombrero blanco. Considere-mos a una de estas dos personas. Ella ve que una de sus colegas tiene sombrero blanco y razona que si tuviera sombrero negro, su colega sería la única que llevara sombrero blanco. De ser así, su colega hubiera comprendido la situación e informado al rey en la primera hora (todas las matemáticas son listas). Que esto no haya sucedido implica que ella también lleva sombrero blanco. En consecuencia, le informa al rey en la segunda hora (cosa que también hace la otra matemática con sombrero blanco pues, de nuevo, todas las matemáticas son listas).

con sombrero blanco. En consecuencia, ella comprendió que debía llevar, también, sombrero blanco. En la hora (k + l)-ésima, ella (junto con todas las otras matemáticas con sombrero blanco) dirán al rey su conclusión.

❐

Ejemplo 1.4

Considere el siguiente juego solitario: para cada entero i, hay un número ilimitado de pelotas marcadas con el número i. Inicialmente nos dan una caja de pelotas y vamos a sacar las pelotas de la caja una por una. Si sacamos una pelota marcada con i, la podemos remplazar por cualquier número finito de pelotas marcadas $1, 2, \ldots, i - 1$ (esto es, no hay remplazo si quitamos una pelota marcada con el 1). El juego termina cuando la caja está vacía. Queremos saber si el juego siempre termina para cualquier caja de pelotas dada inicialmente.

Demostraremos que el juego siempre termina, por medio de inducción sobre n, con el mayor número que aparece en las pelotas de la caja.

1. *Base de la inducción.* Para $n = 1$, hay inicialmente en la caja un número finito de pelotas marcadas con el 1. Como no hay remplazo después de sacar una pelota marcada con el 1, el juego termina después de un número finito de movidas.
2. *Paso de inducción.* Supongamos que el juego termina si el número más grande que aparece en las pelotas es k. Consideremos el caso en que el número más grande que aparece es $k + 1$. De acuerdo con la hipótesis de inducción, al final sacaremos una pelota marcada $k + 1$ (si sólo sacamos pelotas marcadas con $1, 2, \ldots, k$, se terminarán en un número finito de movidas). Repitiendo este argumento, sacaríamos todas las pelotas marcadas con $k + 1$ en un número finito de movidas. De nuevo, por la hipótesis de inducción, el juego termina, a partir de ahí, después de un número finito de movidas.

❐

Ejemplo 1.5

Demostrar por inducción matemática que:

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} \qquad n \geqq 1$$

1. *Base de la inducción.* Para $n = 1$, tenemos

$$1^2 = \frac{1(1+1)(2+1)}{6}$$

2. *Paso de inducción.* Supongamos que

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}$$

Tenemos

$$
\begin{aligned}
1^2 + 2^2 + \cdots + k^2 + (k+1)^2 &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\
&= \frac{(k+1)[k(2k+1) + 6(k+1)]}{6} \\
&= \frac{(k+1)(2k^2 + 7k + 6)}{6} \\
&= \frac{(k+1)(k+2)(2k+3)}{6} \\
&= \frac{(k+1)[(k+1)+1][2(k+1)+1]}{6}
\end{aligned}
$$

❐

Ejemplo 1.6

Demostrar que cualquier entero formado por 3" dígitos idénticos es divisible entre 3" (por ejemplo, 222 y 777 son divisibles entre 3; 222,222,222 y 555,555,555 son divisibles entre 9). Demostraremos el resultado mediante inducción en *n.*

1. *Base de la inducción.* Para $n = 1$, cualquier entero de tres dígitos con los tres dígitos idénticos es divisible entre 3.†
2. *Paso de inducción.* Sea x un entero compuesto de 3^{k+1} dígitos idénticos. Observamos que x se puede escribir como

$$x = y \times z$$

donde y es un entero compuesto por 3^k dígitos idénticos y

$$z = 10^{2 \cdot 3k} + 10^{3k} + 1 = 1\underbrace{000000 \cdots 0}_{3^k - 1 \text{ ceros}}1\underbrace{000000 \cdots 0}_{3^k - 1 \text{ ceros}}1$$

Como suponemos que y es divisible entre 3^k, y z es claramente divisible entre 3, concluimos que x es divisible entre 3^{k+1}.

❐

Ejemplo 1.7

Demostrar que $2^n > n^3$ para $n \geq 10$.

1. *Base de la inducción.* Para $n = 10, 2^{10} = 1024$ que es mayor que 10^3.
2. *Paso de inducción.* Suponga que $2^k > k^3$. Observe que

$$2^{k+1} = 2 \cdot 2^k > \left(1 + \frac{1}{10}\right)^3 \cdot 2^k \geqq \left(1 + \frac{1}{k}\right)^3 \cdot 2^k > \left(1 + \frac{1}{k}\right)^3 \cdot k^3 = (k+1)^3$$

❐

Una forma más "poderosa" del principio de inducción matemática, conocida como *principio de la inducción matemática fuerte,* se puede enunciar así: para una afirmación dada acerca de un número natural *n,* si podemos demostrar que

---

† Recordamos al lector el resultado elemental de que un entero es divisible entre 3 si la suma de sus dígitos es divisible entre 3.

1'. La afirmación es verdadera para $n = n_0$; y

2'. La afirmación es verdadera para $n = k + 1$, suponiendo que la afirmación es verdadera para $n_0 \leqq n \leqq k$,

entonces podemos concluir que la afirmación es verdadera para todos los números naturales $n \geqq n_0$.

En efecto, ésta es una forma más poderosa del principio de inducción matemática presentado al comienzo de esta sección. Específicamente, en el paso de inducción, para demostrar que la afirmación es verdadera para $n = k + 1$, se *nos permite* hacer una suposición más fuerte en 2' (a saber, la afirmación es verdadera para, $n_0 \leqq n \leqq k$) que en la 2 original (a saber, que la afirmación es verdadera para $n — k$). En otras palabras, el principio de la inducción matemática fuerte nos permite llegar a la misma conclusión suponiendo más. Dejamos la demostración del principio de la inducción matemática fuerte para el problema 1.53. Aquí presentamos algunos ejemplos:

Ejemplo 1.8

Un rompecabezas consta de cierto número de piezas. Dos o más piezas con frontera común se pueden juntar y formar una pieza "grande". Para ser más precisos, usamos el término *bloque* para referirnos a una sola pieza o a un cierto número de piezas con fronteras comunes que se pueden juntar para formar una pieza "grande". Así, podemos simplemente decir que se pueden juntar bloques con fronteras comunes para formar otro bloque. Por último, cuando todas las piezas se han colocado juntas en un solo bloque, decimos que hemos armado el rompecabezas. Colocar dos bloques con frontera común se considera una movida. Usaremos el principio de la inducción matemática fuerte para demostrar que para armar un rompecabezas con n piezas se emplean n - 1 movidas.

1. *Base de la inducción.* Un rompecabezas de una pieza no necesita movidas para ser armado.
2. *Paso de inducción.* Supongamos que para cualquier rompecabezas de n piezas, $1 \leq n \leq k$, se emplean n - 1 movidas para armarlo. Ahora consideremos un rompecabezas de $k + 1$ piezas. Para la última movida que termina de armar el rompecabezas, se unen dos bloques -uno con n_1 piezas y el otro con n_2, donde $n_1 + n_2 = k + 1$-para formar uno solo. De acuerdo con la hipótesis de inducción, se emplean n_1 - 1 movidas para juntar un bloque y n_2 - 1 movidas para juntar el otro bloque. Incluida la última movida para unir los dos bloques, el número total de movidas es igual a

$$(n_1 - 1) + (n_2 - 1) + 1 = k + 1 - 1 = k$$

❒

Ejemplo 1.9

Queremos demostrar que cualquier entero positivo n mayor o igual a 2 es primo o es producto de primos.

1. *Base de la inducción.* Para $n = 2$, como 2 es primo, la afirmación es verdadera.
2. *Paso de inducción.* Supongamos que la afirmación es verdadera para cualquier entero n, $2 \leqq n \leqq k$. Para el entero $k + 1$. si $k + 1$ es primo, entonces la afirmación es verdadera. Si $k + 1$ no es primo, entonces se puede escribir como pq, donde $p \leqq k$

y $q \leqq k$. De acuerdo con la hipótesis de inducción, p es primo o producto de primos. Además, q es primo o producto de primos. En consecuencia, pq es producto de primos.

Ejemplo 1.10

Deseamos construir nuestro árbol genealógico para identificar a nuestros antepasados. Con frecuencia los registros familiares incompletos nos impiden ir más atrás de unas cuantas generaciones. La figura 1.5*a* y *b* muestra dos ejemplos de árboles genealógicos. Para simplificar hacemos la hipótesis de que para cualquiera de nuestros antepasados, o podemos ubicar a ambos padres o a ninguno. En un árbol genealógico llamaremos a una persona "hoja" si no podemos ubicar a sus padres, y la llamaremos "nodo interno" si los podemos ubicar.†

Queremos demostrar que en cualquier árbol genealógico el número de hojas es siempre uno o más que el número de nodos internos. Probaremos la afirmación mediante inducción sobre n, el número de personas del árbol genealógico, usando el principio de inducción matemática fuerte.

1. *Base de la inducción.* La afirmación es verdadera cuando $n = 1$. Cuando sólo hay una persona en el árbol genealógico, es una hoja y no hay nodo interno.
2. *Paso de inducción.* Supongamos que la afirmación es verdadera para todos los árboles genealógicos con n personas, para $1 \le n \le k$. Queremos demostrar que la afirmación es verdadera para cualquier árbol genealógico con $k + 1$ personas. Considere el árbol genealógico de un hombre que tiene $k + 1$ personas. Denotemos con;? el número de hojas del árbol y con q el número de nodos internos. Como hay al menos tres personas en el árbol, ambos padres de esta persona están en el árbol.

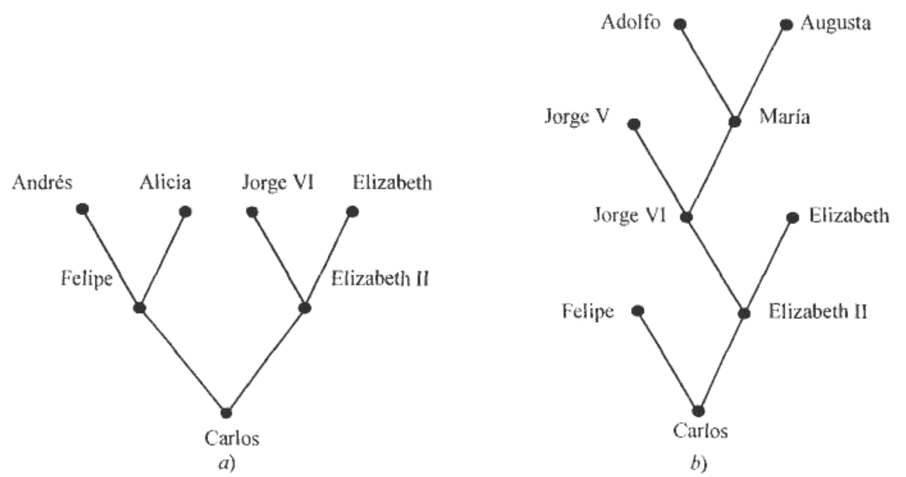


Figura 1.5

† Introducimos estos términos sólo para facilitar nuestra presentación. Se presentarán formalmente en el capítulo 6, cuando estudiemos los árboles como una clase particular de gráficas.

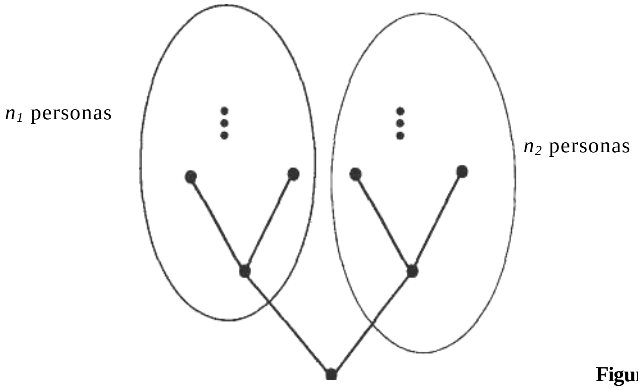


Figura 1.6

Ahora consideremos el árbol genealógico de cada uno de sus padres, como se ilustra en la figura 1.6. Denotemos con n_1 el número de personas del árbol genealógico de su padre, donde p_1 de ellas son hojas y q_1 son nodos internos; con n_2 el número de personas en el árbol genealógico de su madre, donde p_2 son hojas y q_2 son nodos internos. Como $n_1 \leqq k$ y $n_2 \leqq k$, de acuerdo con la hipótesis de inducción

$$p_1 = q_1 + 1$$
$$p_2 = q_2 + 1$$

Como

$$p = p_1 + p_2$$
$$q = q_1 + q_2 + 1$$

tenemos

$$p = q_1 + q_2 + 2 = q + 1$$

❐

1.6 PRINCIPIO DE INCLUSIÓN Y EXCLUSIÓN

En esta sección presentamos algunos resultados sobre la cardinalidad de conjuntos finitos. Para denotar la cardinalidad del conjunto P usaremos $|P|$. Algunos resultados sencillos, cuya deducción se deja al lector, son:

$$
\begin{aligned}
|P \cup Q| &\leqq |P| + |Q| \\
|P \cap Q| &\leqq \text{mín}\,(|P|, |Q|) \\
|P \oplus Q| &= |P| + |Q| - 2|P \cap Q| \\
|P - Q| &\geqq |P| - |Q|
\end{aligned}
$$

A continuación mostramos un resultado menos obvio. Sean $|A_1|$ y $|A_2|$ dos conjuntos; queremos demostrar que

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \tag{1.1}$$

Los conjuntos A_1 *y* A_2 pueden tener algunos elementos en común. Para ser precisos, el número de elementos comunes a A_1 *y* A_2 es $|A_1 \cap A_2|$. Cada uno de estos elementos se cuenta dos veces en $|A_1| + |A_2|$ (una vez en $|A_1|$ y otra en $|A_2|$), aunque debería contarse como un elemento en $|A_1 \cup A_2|$. Por tanto, esa *doble cuenta* de elementos en $|A_1| + |A_2|$ debe ajustarse restando el término $|A_1 \cap A_2|$ en el lado derecho de (1.1). Como ejemplo, supongamos que de un conjunto de 12 libros, 6 son novelas, 7 se publicaron en el año de 1984 y 3 son novelas publicadas en 1984. Denotemos con A_x el conjunto de libros que son novelas y con A_2 el conjunto de los libros publicados en 1984. Tenemos

$$|A_1| = 6 \qquad |A_2| = 7 \qquad |A_1 \cap A_2| = 3$$

En consecuencia, de acuerdo con (1.1),

$$|A_1 \cup A_2| = 6 + 7 - 3 = 10$$

Esto es, hay 10 libros que son novelas o publicaciones de 1984, o ambas. En consecuencia, entre los 12 libros hay 2 que no son novelas y que no fueron publicados en 1984.
Extendiendo el resultado en (1.1) tenemos, para tres conjuntos A_1,A_2 *y* A_3,

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| = {} & |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| \\ & - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3| \end{aligned} \tag{1.2}$$

Como más adelante demostraremos un resultado más general, no demostraremos aquí el resultado (1.2). Por otro lado, sugerimos que el lector verifique el resultado de (1.2) examinando el diagrama de Venn de la figura 1.7. Consideremos algunos ejemplos ilustrativos:

Ejemplo 1.11

Supongamos que tenemos seis computadoras con las siguientes especificaciones:

Computadora	Unidad aritmética de punto flotante	Memoria en disco magnético	Terminal con despliegue gráfico
I	Sf	Sí	No
II	Sí	Sí	Sí
III	No	No	No
IV	No	Sí	Sí
V	No	Sí	No
VI	No	Sí	Sí

Sean A_1,A_2*y* A_3*los* conjuntos de computadoras con unidad aritmética de punto flotante, almacenamiento en disco magnético y terminal con despliegue gráfico, respectivamente. Tenemos

$$|A_1| = 2 \qquad |A_2| = 5 \qquad |A_3| = 3$$

$$|A_1 \cap A_2| = 2 \qquad |A_1 \cap A_3| = 1 \qquad |A_2 \cap A_3| = 3$$

$$|A_1 \cap A_2 \cap A_3| = 1$$

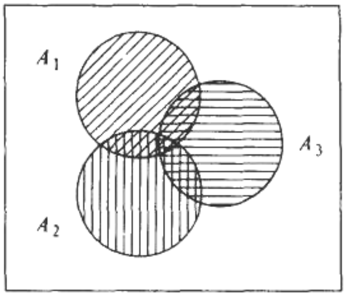


Figura 1.7

En consecuencia,

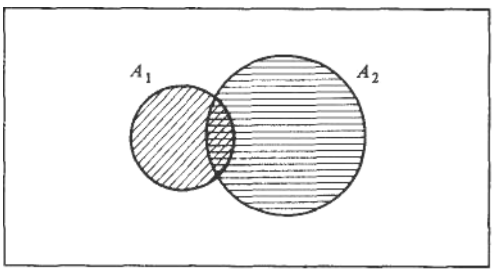
$$|A_1 \cup A_2 \cup A_3| = 2 + 5 + 3 - 2 - 1 - 3 + 1 = 5$$

Esto es, cinco de las seis computadoras tienen uno o más de los tres tipos de *hardware* mencionados.

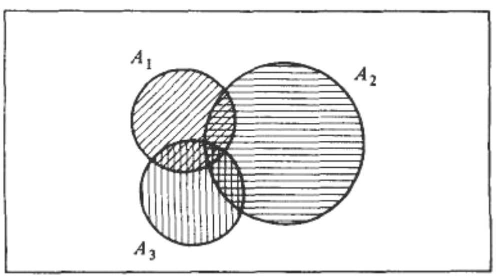
❐

Ejemplo 1.12

De 200 estudiantes, 50 toman el curso de matemáticas discretas, 140 el curso de economía y 24 ambos cursos. Como ambos cursos programaron exámenes para el día siguiente, sólo los estudiantes que no están en ninguno de estos cursos podrán ir **a** la fiesta de la noche anterior. Queremos saber cuántos estudiantes estarán en la fiesta. Al examinar el diagrama de Venn de la figura 1.8*a*, donde A_1 es el conjunto de estudiantes del curso de matemáticas discretas y A_2 es el conjunto de estudiantes del curso de



a)



b)

Figura 1.8

economía, observamos que el número de estudiantes que toman uno o ambos cursos es igual a

$$50+140-24=166$$

En consecuencia, el número de estudiantes que irán a la fiesta es

$$200 - 166 = 34$$

Supongamos que 60 de los 200 son estudiantes de los primeros años. De éstos, 20 toman matemáticas discretas, 45 toman economía y 16 cursan ambas materias. Queremos saber cuántos estudiantes de los últimos años estarán en la fiesta. De acuerdo con el diagrama de Venn de la figura 1.8*b*, donde A_3 es el conjunto de los estudiantes de los primeros años, tenemos

$$|A_1 \cup A_2 \cup A_3| = 50 + 140 + 60 - 24 - 20 - 45 + 16 = 177$$

Así, el número de estudiantes de los últimos años que irán a la fiesta es

$$200 - 177 = 23$$

❐

Ejemplo 1.13 En una fábrica se ensamblaron treinta autos. Las opciones disponibles fueron: radio, aire acondicionado y llantas de cara blanca. Sabemos que 15 autos tienen radio, 8 aire acondicionado y 6 tienen llantas de cara blanca. Más aún 3, tienen las tres opciones. Sean A_1, A_2, A_3 los conjuntos de autos con radio, aire acondicionado y llantas de cara blanca, respectivamente. Como

$$|A_1| = 15 \qquad |A_2| = 8 \qquad |A_3| = 6$$

y

$$|A_1 \cap A_2 \cap A_3| = 3$$

de acuerdo con (1.2)

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= 15 + 8 + 6 - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + 3 \\ &= 32 - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| \end{aligned}$$

Como

$$\begin{aligned} |A_1 \cap A_2| &\geqq |A_1 \cap A_2 \cap A_3| \\ |A_1 \cap A_3| &\geqq |A_1 \cap A_2 \cap A_3| \\ |A_2 \cap A_3| &\geqq |A_1 \cap A_2 \cap A_3| \end{aligned}$$

tenemos

$$|A_1 \cup A_2 \cup A_3| \leqq 32 - 3 - 3 - 3 = 23$$

Esto es, hay *a lo más* 23 autos que tienen una o más opciones. En consecuencia, hay *al menos* 7 autos que no tienen ninguna opción.

❐

En el caso general, para los conjuntos $A_1, A_2, \ldots, A_r$, tenemos

$$|A_1 \cup A_2 \cup \cdots \cup A_r| = \sum_i |A_i| - \sum_{1 \leq i < j \leq r} |A_i \cap A_j| + \sum_{1 \leq i < j < k \leq r} |A_i \cap A_j \cap A_k| + \cdots + (-1)^{r-1} |A_1 \cap A_2 \cap \cdots \cap A_r| \qquad (1.3)$$

Aunque no es difícil visualizar el resultado en (1.2), el resultado en (1.3) no es tan obvio. Demostraremos (1.3) mediante inducción sobre el número de conjuntos r. Es claro que (1.1) nos puede servir como base de la inducción. Como paso de inducción suponemos que (1.3) es válido para cualesquiera r - 1 conjuntos. Señalemos primero que, viendo $(A_1 \cup A_2 \cup \cdots \cup A_{r-1})$ y A_r como dos conjuntos, de acuerdo con (1.1) tenemos

$$|A_1 \cup A_2 \cup \cdots \cup A_r| = |A_1 \cup A_2 \cup \cdots \cup A_{r-1}| + |A_r| - |A_r \cap (A_1 \cup A_2 \cup \cdots \cup A_{r-1})| \tag{1.4}$$

Ahora,

$$|A_r \cap (A_1 \cup A_2 \cup \cdots \cup A_{r-1})| = |(A_r \cap A_1) \cup (A_r \cap A_2) \cup \cdots \cup (A_r \cap A_{r-1})|$$

De acuerdo con la hipótesis de inducción, para los r - 1 conjuntos $A_r \cap A_1, A_r \cap A_2, \ldots, A_r \cap A_{r-1}$, tenemos

$$\begin{aligned}
|(A_r \cap A_1) \cup (A_r \cap A_2) \cup \cdots \cup (A_r \cap A_{r-1})| \\
= {} & |A_r \cap A_1| + |A_r \cap A_2| + \cdots + |A_r \cap A_{r-1}| \\
& - (A_r \cap A_1) \cap (A_r \cap A_2)| - |(A_r \cap A_1) \cap (A_r \cap A_3)| \\
& - \cdots \\
& + |(A_r \cap A_1) \cap (A_r \cap A_2) \cap (A_r \cap A_3) + \cdots \\
& - \cdots \\
& + (-1)^{r-2} |(A_r \cap A_1) \cap (A_r \cap A_2) \cap \cdots \cap (A_r \cap A_{r-1})| \\
= {} & |A_r \cap A_1| + |A_r \cap A_2| + \cdots + |A_r \cap A_{r-1}| \\
& - |A_r \cap A_1 \cap A_2| - |A_r \cap A_1 \cap A_3| - \cdots \\
& + |A_r \cap A_1 \cap A_2 \cap A_3| + \cdots \\
& - \cdots \\
& + (-1)^{r-2} |A_r \cap A_1 \cap A_2 \cap \cdots \cap A_{r-1}|
\end{aligned} \tag{1.5}$$

Además, de acuerdo con la hipótesis de inducción, para los r - 1 conjuntos $A_1, A_2, \ldots, A_{r-1}$, tenemos

$$\begin{aligned}
|(A_1 \cup A_2 \cup \cdots \cup A_{r-1}| = {} & |A_1| + |A_2| + \cdots \\
& - |A_1 \cap A_2| - |A_1 \cap A_3| - \cdots \\
& + \cdots \\
& + (-1)^{r-2} A_1 \cap A_2 \cap \cdots \cap A_{r-1}|
\end{aligned} \tag{1.6}$$

Sustituyendo (1.5) y (1.6) en (1.4), obtenemos (1.3).

Ejemplo 1.14 Determinemos el número de enteros entre 1 y 250 que son divisibles entre cualquiera de los enteros 2, 3, 5 y 7. Denotemos con A_1 el conjunto de los enteros entre 1 y 250 que son divisibles entre 2, con A_2 el conjunto de los enteros divisibles entre 3, con A_3 el conjunto de los enteros divisibles entre 5 y con A_4 el conjunto de los enteros divisibles entre 7. Como

$$|A_1| = \left\lfloor \frac{250}{2} \right\rfloor^{\dagger} = 125 \qquad |A_2| = \left\lfloor \frac{250}{3} \right\rfloor = 83$$

$$|A_3| = \left\lfloor \frac{250}{5} \right\rfloor = 50 \qquad |A_4| = \left\lfloor \frac{250}{7} \right\rfloor = 35$$

$$|A_1 \cap A_2| = \left\lfloor \frac{250}{2 \times 3} \right\rfloor = 41 \qquad |A_1 \cap A_3| = \left\lfloor \frac{250}{2 \times 5} \right\rfloor = 25$$

$$|A_1 \cap A_4| = \left\lfloor \frac{250}{2 \times 7} \right\rfloor = 17 \qquad |A_2 \cap A_3| = \left\lfloor \frac{250}{3 \times 5} \right\rfloor = 16$$

$$|A_2 \cap A_4| = \left\lfloor \frac{250}{3 \times 7} \right\rfloor = 11 \qquad |A_3 \cap A_4| = \left\lfloor \frac{250}{5 \times 7} \right\rfloor = 7$$

$$|A_1 \cap A_2 \cap A_3| = \left\lfloor \frac{250}{2 \times 3 \times 5} \right\rfloor = 8 \qquad |A_1 \cap A_2 \cap A_4| = \left\lfloor \frac{250}{2 \times 3 \times 7} \right\rfloor = 5$$

$$|A_1 \cap A_3 \cap A_4| = \left\lfloor \frac{250}{2 \times 5 \times 7} \right\rfloor = 3 \qquad |A_2 \cap A_3 \cap A_4| = \left\lfloor \frac{250}{3 \times 5 \times 7} \right\rfloor = 2$$

$$|A_1 \cap A_2 \cap A_3 \cap A_4| = \left\lfloor \frac{250}{2 \times 3 \times 5 \times 7} \right\rfloor = 1$$

tenemos

$$\begin{aligned} |A_1 \cup A_2 \cup A_3 \cup A_4| &= 125 + 83 + 50 + 35 - 41 - 25 - 17 - 16 - 11 - 7 \\ &\quad + 8 + 5 + 3 + 2 - 1 = 193 \end{aligned}$$

❐

En capítulos posteriores presentaremos más ejemplos de aplicación de la fórmula (1.3), que se conoce como *principio de inclusión y exclusión.*

* 1.7 MULTICONJUNTOS

Un conjunto es una colección de objetos distintos, sin embargo, hay ocasiones en que encontramos colecciones de objetos no distintos. Por ejemplo, considere los nombres de los estudiantes de una clase. Podemos tener dos o más estudiantes con el mismo nombre, y quizá

† Usamos |x| para denotar el mayor entero que es menor o igual que *x*.

nos interese hablar sobre la colección de los nombres de los estudiantes. Definimos un *multiconjunto* como una colección de objetos que no son necesariamente distintos. Así $\{a, a, a, b, b, c\}$,$\{a, a, a, a\}$, $\{a, b, c)$ y $\{\ \}$ son ejemplos de multiconjuntos. La *multiplicidad* de un elemento en un multiconjunto se define como el número de veces que el elemento aparece en el multiconjunto. Así, la multiplicidad del elemento a en el multiconjunto $\{a, a, a, c, d, d)$ es 3. La multiplicidad del elemento b es 0, la multiplicidad del elemento c es 1, y la multiplicidad del elemento d es *2.* Observe que los conjuntos son sólo casos particulares de los multiconjuntos, donde la multiplicidad de un elemento es 0 o 1. La cardinalidad de un multiconjunto se define como la cardinalidad que le corresponde *suponiendo* que los elementos del multiconjunto fueran distintos.

Sean P y Q dos multiconjuntos. La unión de P y Q, denotada con $P \cup Q$, es un multiconjunto tal que la multiplicidad de un elemento en $P \cup Q$ es igual al máximo de las multiplicidades del elemento en P y en Q. Así, para $P = \{a, a, a, c, d, d\}$ y $Q = \{a, a, b, c, c)$

$$P \cup Q = \{a, a, a, b, c, c, d, d\}$$

Por ejemplo, sea el multiconjunto R = {ingeniero eléctrico, ingeniero eléctrico, ingeniero eléctrico, ingeniero mecánico, matemático, matemático, físico} el personal requerido para la primera fase de un proyecto de ingeniería, y el multiconjunto S = {ingeniero eléctrico, ingeniero mecánico, ingeniero mecánico, matemático, científico en computación, científico en computación} el personal requerido para la segunda fase del proyecto. El multiconjunto $R \cup S$ es el personal que se deberá contratar para el proyecto.

La intersección de P y Q, denotada con $P \cap Q$, es un multiconjunto tal que la multiplicidad de un elemento en P n Q es igual al mínimo de las multiplicidades del elemento en P y en Q. Así, para $P = \{a, a,a,c,d,d\}$ y $Q= \{a, a, b, c, c\}$,

$$P \cap Q = \{a, a, c\}$$

Para el ejemplo del proyecto de ingeniería, el multiconjunto $R \cap S$ es el personal que participará en ambas fases del proyecto.

La diferencia de P y Q, denotada con $P - Q$, es un multiconjunto tal que la multiplicidad de un elemento en $P - Q$ es igual a la multiplicidad del elemento en P menos la multiplicidad del elemento en Q si la diferencia es positiva, y es igual a 0 si la diferencia es 0 o negativa. Por ejemplo, sea $P = \{a, a, a, b, b, c, d, d, e\}$ y $Q = \{a, a, b, b, b, c, c, d,d,f\}$. Tenemos

$$P-Q=\{a,e\}$$

Para este ejemplo, el multiconjunto $R - S$ es el personal que será reasignado después de la primera fase del proyecto.

Las definiciones de unión, intersección y diferencia de multiconjuntos se han escogido de manera que sean consistentes con las de conjuntos. Aquí no definimos la diferencia simétrica de dos multiconjuntos, pero el lector interesado puede remitirse al problema 1.64.

Por último, definimos la suma de dos multiconjuntos P y Q, denotada con $P + Q$, como el multiconjunto tal que la multiplicidad de un elemento de $P + Q$ es igual a la suma de las multiplicidades del elemento en P y en Q. Observe que no hay definición correspondiente para la suma de dos conjuntos. Por ejemplo, sean $P = \{a, a, b, c, c\}$ y $Q = \{a, b, b,d\}$.

Tenemos que $P + Q=\{a,a, a, b, b, b, c,c,d)$. Otro ejemplo: sea$R$ el multiconjunto que contiene los números de cuenta de todas las transacciones de un banco en un cierto día, y S el multiconjunto que contiene los números de cuenta de todas las transacciones en el día siguiente. R *y* S son multiconjuntos, porque un número de cuenta pudo tener más de una transacción en un día. Así, $R + S$ es un registro combinado de los números de cuenta de las transacciones en estos dos días.

1.8 PROPOSICIONES

Una proposición es una frase declarativa que es verdadera o falsa. "Llovió ayer", "la presión dentro de la cámara del reactor sobrepasa el límite de seguridad" y "habrá pollo para la cena" son ejemplos de proposiciones. Por otro lado, "¿qué hora es?" y "por favor, mande su informe a la brevedad posible" no son proposiciones, pues no son frases declarativas y, en consecuencia, no tiene sentido referirse a ellas como verdaderas o falsas. No descartamos la posibilidad de que una proposición sea definitivamente verdadera, como " 15 es divisible entre 3", ni la posibilidad de que una proposición sea definitivamente falsa, como "Champaign es la capital de Illinois". Una proposición que es verdadera en cualquier circunstancia se conoce como *tautología, y* una proposición que es falsa en cualquier circunstancia se conoce como *contradicción.*

Con frecuencia nos referiremos a proposiciones mediante nombres simbólicos. Por ejemplo, *sip* denota la proposición "todos los estudiantes de la clase aprobaron el examen final", podemos decir que p es verdadera o p es falsa dependiendo de los resultados del examen final. Una proposición tiene dos posibilidades, ser verdadera o falsa y a éstas se les conoce como los dos valores posibles que puede tomar una proposición. Se acostumbra usar una Fpara designar el valor *verdadero y F,* para el valor *falso.* En consecuencia, en lugar de decir que la proposición "todos los estudiantes de la clase aprobaron el examen final" es verdadera, podemos simplemente decir que el valor de p es V.

Se dice que dos proposiciones p *y* q son *equivalentes* si cuando p es verdadera también q es verdadera, cuando p es falsa también q es falsa, y viceversa. Por ejemplo, las dos proposiciones "el agua se congeló esta mañana" y "la temperatura estuvo debajo de 0°C esta mañana" son equivalentes. También las proposiciones "él nació en 1934" y "él tendrá 60 años en 1994", son equivalentes. Las proposiciones "iré al juego de pelota esta noche" y "no hay clases mañana" pueden o no ser equivalentes. Por otro lado, las dos proposiciones "x es un número primo" y "x no es divisible entre 2" no son equivalentes, debido a que si x no es divisible entre 2 no significa necesariamente que x sea un número primo.

Se pueden combinar proposiciones para construir nuevas proposiciones. Por ejemplo, respecto a la operación de una compañía, denotemos con p la proposición de que el volumen de ventas mensuales es menor que \$ 200 000 y con q la proposición de que el gasto mensual excede \$ 200 000. Quizá nos interese construir una proposición que describa la situación de que el volumen de ventas es menor que \$ 200 000 y que el gasto mensual excede \$ 200 000. También puede interesarnos construir una proposición que describa la situación en la cual el volumen de ventas mensuales sea menor que \$ 200 000 o el gasto mensual exceda \$ 200 000. Es obvio que estas proposiciones son ejemplos de "combinaciones" de las proposiciones p y q.

p	q	$p \vee q$
F	F	F
F	V	V
V	F	,V
V	V	V

p	q	$p \wedge q$
F	F	F
F	V	F
V	F	F
V	V	V

p	$\overline{p}$
F	V
V	F

Figura 1.9

Sean p y q dos proposiciones. Definimos la *disyunción* de p y q, denotada con $p \vee q$, como la proposición que es verdadera cuando ya sea que p o q o ambas sean verdaderas, y es falsa cuando ambas, p y q, sean falsas. En el ejemplo de la operación de una compañía, la disyunción de la proposición de que el volumen de ventas mensuales es menor que \$ 200 000 y la proposición de que el gasto mensual excede \$ 200 000 es la proposición de que o el volumen de ventas mensuales es menor que \$ 200 000 o el gasto mensual excede \$ 200 000 o ambas cosas.

Sean p y q dos proposiciones. Definimos la *conjunción* de p y q, denotada con $p \wedge q$, como la proposición que es verdadera cuando ambas, p y q, sean verdaderas, y es falsa, cuando p o q, o ambas, sean falsas. En el ejemplo tratado acerca de la operación de una compañía, la conjunción de la proposición de que el volumen de ventas mensuales es menor que \$ 200 000 y la proposición de que el gasto mensual excede \$ 200 000, es la proposición de que tanto el volumen de ventas mensuales es menor que \$ 200 000 como el gasto mensual excede \$ 200 000.

Sea p una proposición. Definimos la *negación* de p, denotada por $\overline{p}$,† como la proposición que es verdadera cuando p es falsa, y es falsa cuando p es verdadera. Así, la negación de la proposición de que el volumen de ventas mensuales es menor que \$ 200 000 es la proposición de que el volumen de ventas mensuales excede o iguala \$ 200 000.

Una proposición obtenida de la combinación de otras proposiciones se conoce como proposición *compuesta.* Una proposición que no es combinación de otras se conoce como proposición *atómica.* En otras palabras, una proposición compuesta está formada de proposiciones atómicas. Una manera conveniente y precisa de describir la definición de una proposición compuesta es una tabla como la mostrada en la figura 1.9, donde los valores de una proposición compuesta están especificados para todas las posibilidades de los valores de las proposiciones atómicas de la proposición compuesta. En concreto, las tablas de la figura 1.9 muestran las definiciones de la disyunción y conjunción de dos proposiciones, y de la negación de una proposición. Dichas tablas se llaman *tablas de verdad* de las proposiciones compuestas.

† También se usa la notación -p.

p	q	$p \to q$
F	F	V
F	V	V
V	F	F
V	V	V

Figura 1.10

Debemos tener presentes otras dos maneras importantes de construir proposiciones compuestas: considere las proposiciones "la temperatura rebasó los 70°C" y "la alarma sonará", que denotaremos con p y q, respectivamente. Además, consideremos la proposición "si la temperatura rebasa los 70°C, entonces la alarma sonará", que denotaremos con r. Es fácil ver que r es verdadera si la alarma suena cuando la temperatura rebasa los 70°C (tanto p como q son verdaderas), y r es falsa si la alarma no suena cuando la temperatura rebasa los 70°C *(p* es verdadera y q es falsa). Por otro lado, cuando la temperatura es igual o menor que 70°C *(p* es falsa), la proposición *r no puede ser falsa,* sin importar si la alarma suena o no. En consecuencia, decimos que r siempre es verdadera cuando la temperatura es igual o menor que 70°C.

Formalicemos el concepto de combinar dos proposiciones p y q para formar una que diga "si p entonces *q"* como la introducida en el ejemplo anterior. Sean p y q dos proposiciones. Definimos la proposición "si p entonces q", denotada *con* $p \to q$, que es verdadera si tanto/* como q son verdaderas o si*p* es falsa, y que es falsa *úp* es verdadera y q es falsa, como se indica en la tabla de verdad de la figura 1.10. La proposición compuesta "si p entonces *q"* también se lee "*p* implica *q".*

Un lector que lea por vez primera la proposición compuesta "si p entonces *q"* quizá tenga alguna duda acerca del hecho de que la afirmación compuesta sea verdadera cuando p sea falsa, sin importar que q sea verdadera o no. Examinemos otros ejemplos. Al considerar la afirmación "si lo intentas, lo lograrás", es claro que si lo intentas y lo logras, la afirmación es cierta. Si lo intentas y fallas, la afirmación es falsa. Sin embargo, si no lo intentas, no hay manera de argumentar que la afirmación es falsa. Como no ser falsa significa que es verdadera, podemos concluir que si no lo intentas entonces la afirmación es verdadera. Otro ejemplo: considere la orden del oficial de seguridad de una compañía de que todo visitante debe portar gafete. La orden se puede reformular como una proposición "si alguien es visitante, entonces debe portar gafete". Para verificar que se ha acatado la orden (que la proposición es verdadera), debemos revisar a cada persona en las oficinas, una por una. Si se trata del visitante, podemos determinar que la orden se ha acatado al ver si porta gafete. Por otro lado, si no es visitante, entonces no hay manera de concluir que la orden no se ha acatado y, por tanto, la afirmación es verdadera. Como un ejemplo más, recordemos que en la sección 1.1 hicimos la afirmación de que el conjunto vacío es subconjunto de cualquier conjunto. De acuerdo con la definición de subconjunto de un conjunto, esta afirmación puede reformularse como "si x es un elemento del conjunto vacío, entonces x es un elemento de cualquier conjunto". Es obvio que la afirmación es verdadera.

p		$p \leftrightarrow q$
F	F	V
F	V	F
V	F	F
V	V	V

Figura 1.11

Ejemplo 1.15

Juan hizo las siguientes afirmaciones:

1. Yo amo a Luisa.
2. Si yo amo a Luisa, entonces también amo a Verónica.

Dado que Juan dijo la verdad o mintió en ambos casos, determine si en verdad Juan ama a Luisa. Suponga que Juan mintió. Entonces, de acuerdo con la afirmación 1, él no ama a Luisa. Se sigue que la afirmación 2 debe ser verdadera, lo cual es una contradicción. En consecuencia, Juan debe haber dicho la verdad y podemos confirmar que en realidad ama a Luisa.

Se puede ser un poco más formal y hacer que p denote la afirmación "Juan ama a Luisa", y q denote la afirmación "Juan ama a Verónica". La tabla de verdad de la figura 1.10 muestra que es posible que tanto p como $p \rightarrow q$ sean verdaderas, pero que no es posible que ambas sean falsas.

□

Denotemos con p la proposición "compraremos una computadora nueva" y con q la proposición "se dispone de fondos adicionales". Considere la proposición "se comprará una computadora nueva si y sólo si se dispone de fondos adicionales", que denotaremos con r. Como se ve, r es verdadera si, en efecto, se compra una computadora nueva cuando se disponga de fondos adicionales (tanto p como q son verdaderas). La proposición r también es verdadera si no se compra una computadora nueva cuando no se disponga de fondos adicionales (tanto p como q son falsas). Por otro lado, r es falsa si se compra una computadora nueva aunque no se disponga de fondos adicionales (p es verdadera y q es falsa), o si no se compra una computadora nueva aunque se disponga de fondos adicionales (p es falsa y q es verdadera).

Sean p y q dos proposiciones. Definimos la proposición "p si y sólo si q", denotada con $p \leftrightarrow q$, que es verdadera si ambas, p y q, son verdaderas o si ambas, p y q, son falsas, y que es falsa si p es verdadera cuando q es falsa y si p es falsa cuando q es verdadera. La tabla de verdad de la figura 1.11 muestra la definición de $p \leftrightarrow q$.

Ejemplo 1.16

En una isla hay dos tribus de nativos. Todos los nativos de la primera tribu siempre dicen la verdad, mientras que todos los nativos de la otra tribu siempre mienten. Al llegar a la isla y preguntar a un nativo si hay oro en la isla, él responde, "hay oro en la isla si y sólo si yo siempre digo la verdad". ¿De qué tribu es? ¿Hay oro en la isla? Resulta que no podemos determinar de qué tribu es, pero podemos determinar si hay oro en la isla. Designemos con p la proposición de que él siempre dice la verdad y con q la proposición de que hay oro en la isla. Así, su respuesta es la proposición $p \leftrightarrow q$. Supongamos que siempre dice la verdad, esto es, que la proposición q es verdadera. Más aún, su respuesta

p	q	$p \wedge q$	$\overline{p} \wedge \overline{q}$	$(p \wedge q) \vee (\overline{p} \wedge \overline{q})$	$((p \wedge q) \wedge (\overline{p} \wedge \overline{q})) \to p)$
F	F	F	V	V	F
F	V	F	F	F	V
V	F	F	F	F	V
V	V	V	F	V	V

Figura 1.12

debe ser verdadera, esto es, $p \leftrightarrow q$ es verdadera. En consecuencia, q debe ser verdadera. Supongamos que siempre miente, esto es, la proposición p es falsa. Además, su respuesta es una mentira, lo cual significa que $p \leftrightarrow q$ es falsa. En consecuencia, q debe ser verdadera. Entonces, en ambos casos podemos concluir que hay oro en la isla, sin importar de qué tribu sea el nativo.

□

Para construir nuevas proposiciones podemos combinar proposiciones compuestas. Por ejemplo, sean p y q proposiciones. Tenemos $p \wedge q$ y $(\overline{p} \wedge \overline{q})$, en consecuencia,

$$(p \wedge q) \vee (\overline{p} \wedge \overline{q}) \to p \tag{1.7}$$

como proposiciones compuestas en las que se han usado los paréntesis como delimitadores. El valor de una proposición compuesta siempre se puede determinar construyendo su tabla de verdad paso a paso. Como se muestra en la tabla de la figura 1.12, los registros de las columnas que corresponden a las proposiciones compuestas se pueden construir columna a columna de izquierda a derecha.

Podemos determinar si dos proposiciones son equivalentes al examinar sus tablas de verdad. Por ejemplo, de acuerdo con las tablas de verdad de las figuras 1.9 y 1.12, la proposición de (1.7) es equivalente a la proposición $p \vee q$. Como otro ejemplo, las dos proposiciones $p \leftrightarrow q$ y $(p \to q) \wedge (q \to p)$ son equivalentes al comparar las tablas de verdad de las figuras 1.11 y 1.13. En efecto, vemos que no fue accidental escoger la notación $p \leftrightarrow q$.

Ejemplo 1.17

Hay dos restaurantes, uno junto a otro. Uno tiene un letrero que dice "la buena comida no es barata" y el otro tiene un letrero que dice "la comida barata no es buena". ¿Dicen lo mismo? Denotemos con g la proposición de que la comida es buena y con c la proposición de que la comida es barata. El primer letrero se puede escribir como

p	q	$p \to q$	$q \to p$	$(p \to q) \wedge (q \to p)$
F	F	V	V	V
F	V	V	F	F
V	F	F	V	F
V	V	V	V	V

Figura 1.13

p	q	$\bar{g}$	$\bar{c}$	$g \to \bar{c}$	$c \to \bar{g}$
F	F	V	V	V	V
F	V	V	F	V	V
V	F	F	V	V	V
V	V	F	F	F	F

Figura 1.14

$g \to \bar{c}$ y el segundo como $c \to \bar{g}$. La tabla de verdad de la figura 1.14 muestra que, en efecto, los dos letreros dicen lo mismo. ❐

Ejemplo 1.18 Como ejemplo final, pedimos al lector verificar que las dos proposiciones

$$((p \wedge q) \vee (p \wedge r)) \to s$$

y

$$((\bar{p} \vee (\bar{q} \wedge \bar{r})) \vee s$$

son equivalentes. ❐

En el capítulo 12 estudiaremos el tema de construir, manipular y simplificar proposiciones compuestas, después de desarrollar el concepto de álgebras booleanas.

Es probable que algún lector reconozca que la combinación de proposiciones para construir nuevas proposiciones tiene una fuerte analogía con el hecho de que las combinaciones de conjuntos construyen nuevos conjuntos. Para ser precisos, notamos la analogía entre los conceptos de unión de dos conjuntos y de disyunción de dos proposiciones, entre los conceptos de intersección de dos conjuntos y de conjunción de dos proposiciones, y entre los conceptos de complemento de un conjunto y la negación de una proposición. Dicha analogía no es accidental, como veremos en el capítulo 12.

1.9 NOTAS Y REFERENCIAS

Se pueden usar varios libros como referencia general para éste. Véase, por ejemplo, Arbib, Kfoury y Molí [1], Berztiss [2], Birkhoff y Bartee [3], Bogart [4], Cohén [5], Gilí [6], Kemeny, Snell y Thompson [10], Knuth [11], Kolman y Busby [12], Korfhage [13], Levy [14], Liu [15], Prather [17], Preparatay Yeh [18], Sahni [19], Stanaty McAllister [22], Stone [24], Tremblay y Manohar [26] y Tucker [27]. Para profundizar en la teoría de conjuntos, véase Halmos [9] y Stoll [23]. Pólya [16], Sominskii [21] y Golovina [8] son tres libros muy agradables que tratan algo más que el tema de inducción matemática. Para un mayor estudio del principio de inclusión y exclusión véase el capítulo 4 de Liu [15]. Suppes [25] es una referencia útil para el material de la sección 1.8 sobre proposiciones. Véase también el fascinante libro de Smullyan [20].

1. Arbib, M. A., A. J. Kfoury y R. N. Molí: *A Basisfor Theoretical Computer Science,* Springer-Verlag, Nueva York, 1981.
2. Berztiss,A. T.: *Data Structures: Theory and Practice,* 2ª. ed., Academic Press, Nueva York, 1975.
3. Birkhoff, G. y T. C. Bartee: *Modern Applied Algebra,* McGraw-Hill BookCompany, Nueva York, 1970.
4. Bogart, K. R: *Introductory Combinatorics,* Pitman Publishing, Marshfield, Mass., 1983.
5. Cohen, I. A. C: *Basic Techniques of Combinatorial Theory,* John Wiley & Sons, Nueva York, 1978.
6. Gill, A.: *Applied Algebra for the Computer Sciences,* Prentice-Hall, Englewood Cliffs, N. J., 1976.
7. Golomb, S. W.: *Polyominoes,* Scribner's, Nueva York, 1965.
8. Golovina, L. I. e I. M. Yaglom: *Induction in Geometry,* D. C. Heath and Company, Boston, 1963.
9. Halmos, R: *Naive Set Theory,* D. Van Nostrand Company, Princeton, N. J., 1960.
10. Kemeny, J. G., J. L. Snell y G. L. Thompson: *Introduction to Finite Mathematics,* 2ª. ed., Prentice-Hall, Englewood Cliffs, N. J., 1966.
11. Knuth, D. E.: *The Art of Computer Programming, vol. 1, Fundamental Algorithms,* 2ª. ed., Addison-Wesley Publishing Company, Reading, Mass., 1973.
12. Kolman, B. y R. C. Busby: *Discrete Mathematical Structures for Computer Science,* Prentice-Hall, Englewood Cliffs, N. J., 1984.
13. Korfhage, R. R.: *Discrete Computational Structures,* Academic Press, Nueva York, 1974.
14. Levy, L. S.: *Discrete Structures of Computer Science,* John Wiley & Sons, Nueva York, 1980.
15. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
16. Pólya, G.: *Induction and Analogy in Mathematics,* Princeton University Press, Princeton, N. J., 1954.
17. Prather, R. E.: *Discrete Mathematical Structures for Computer Science,* Houghton Mifflin Company, Boston, 1976.
18. Preparata, F. P. y R. T. Yeh: *Introduction to Discrete Structures,* Addison-Wesley Publishing Company, Reading, Mass., 1973.
19. Sahni, S.: *Concepts in Discrete Mathematics,* Camelot Press, Fridley, Minn., 1981.
20. Smullyan, R.: *What Is the Name of This Book—The Riddle of Dracula and Other Logical Puzzles,* Prentice-Hall, Englewood Cliffs, N. J., 1978.
21. Sominskii, I. S.: *The Method of Mathematical Induction,* D. C. Heath and Company, Boston, 1963.
22. Stanat, D. F. y D. F. McAllister: *Discrete Mathematics in Computer Science,* Prentice-Hall, Englewood Cliffs, N. J., 1977.
23. Stoll, R. R.: *Set Theory and Logic,* W. H. Freeman and Company, San Francisco, 1963.
24. Stone, H. S.: *Discrete Mathematical Structures and Their Applications,* Science Research Associates, Palo Alto, California, 1973.
25. Suppes, R: *Introduction to Logic,* D. Van Nostrand Company, Princeton, N. J., 1957.
26. Tremblay, J. P. y R. P. Manohar: *Discrete Mathematical Structures with Applications to Computer Science,* McGraw-Hill Book Company, Nueva York, 1975.
27. Tucker, A.: *Applied Combinatorics,* John Wiley & Sons, Nueva York, 1980.

PROBLEMAS

1.1 Determine si las siguientes afirmaciones son verdaderas o falsas. Explique brevemente la respuesta.

a) $\varnothing \subseteq \varnothing$
b) $\varnothing \in \varnothing$

c) $\varnothing \subseteq \{\varnothing\}$
d) $\varnothing \in \{\varnothing\}$
e) $\{\varnothing\} \subseteq \varnothing$
f) $\{\varnothing\} \in \varnothing$
g) $\{\varnothing\} \subseteq \{\varnothing\}$
h) $\{\varnothing\} \in \{\varnothing\}$
i) $\{a, b\} \subseteq \{a, b, c, \{a, b, c\}\}$
j) $\{a, b\} \in \{a, b, c, \{a, b, c\}\}$
k) $\{a, b\} \subseteq \{a, b, \{\{a, b\}\}\}$
l) $\{a, b\} \in \{a, b, \{\{a, b\}\}\}$
m) $\{a, \varnothing\} \subseteq \{a, \{a, \varnothing\}\}$
n) $\{a, \varnothing\} \in \{a, \{a, \varnothing\}\}$

1.2 Diga cuáles son los conjuntos siguientes:
a) $\varnothing \cup \{\varnothing\}$
b) $\varnothing \cap \{\varnothing\}$
c) $\{\varnothing\} \cup \{a, \varnothing, \{\varnothing\}\}$
d) $\{\varnothing\} \cap \{a, \varnothing, \{\varnothing\}\}$
e) $\varnothing \oplus \{a, \varnothing, \{\varnothing\}\}$
f) $\{\varnothing\} \oplus \{a, \varnothing, \{\varnothing\}\}$

1.3 *a)* Sean *A y B* conjuntos tales que $(A \cup B) \subseteq B$ y $B \subsetneq A$. Dibuje los diagramas de Venn correspondientes.
b) Sean *A, B* y *C* conjuntos tales que $A \subseteq B, A \subseteq C, (B \cap C) \subseteq A$ y $A \subseteq (B \cap C)$. Dibuje los diagramas de Venn correspondientes.
c) Sean *A, B y C* conjuntos tales que $(A \cap B \cap C) = \varnothing, (A \cap B) \neq \varnothing, (A \cap C) \neq \varnothing$ y $(B \cap C) \neq \varnothing$. Dibuje los diagramas de Venn correspondientes.

1.4 Dé un ejemplo de conjuntos *A, B y C* tales que $A \in B, B \in C$, y $A \notin C$.

1.5 Determine cuáles de las siguientes afirmaciones acerca de conjuntos arbitrarios *A, B y C* son verdaderas. Justifique la respuesta.
a) Si $A \in B$ y $B \subseteq C$, entonces $A \in C$.
b) Si $A \in B$ y $B \subseteq C$, entonces $A \subseteq C$.
c) Si $A \subseteq B$ y $B \in C$, entonces $A \in C$.
d) Si $A \subseteq B$ y $B \in C$, entonces $A \subseteq C$.

1.6 Sean *A, B y C* subconjuntos de *U.* Puesto que

$$A \cap B = A \cap C$$
$$\overline{A} \cap B = \overline{A} \cap C$$

¿es necesario que *B = C1* Justifique la respuesta.

1.7 Ya que

$$(A \cap C) \subseteq (B \cap C)$$
$$A \cap \overline{C} \subseteq (B \cap \overline{C})$$

demostrar que $A \subseteq B$.

1.8 ¿Qué se puede decir acerca de los conjuntos *P y Q* si
a) $P \cap Q = P$?
b) $P \cup Q = P$?
c) $P \oplus Q = P$?
d) $P \cap Q = P \cup Q$?

1.9 *a)* Sean $A \subseteq B$ y $C \subseteq D$ ¿Siempre sucede que $(A \cup C) \subseteq (B \cup D)$? ¿Siempre sucede que $(A \cap C) \subseteq (B \cap D)$?
b) Sean $W \subset X$ y $Y \subset Z$ ¿Siempre sucede que $(W \cup Y) \subset (X \subset Z)$? ¿Siempre sucede que $(W \cap Y) \subset (X \cap Z)$?

1.10 *a)* Puesto que $A \cup B = A \cup C$, ¿es necesario que B = C7
b) Puesto que $A \cap B = A \cap C$, ¿es necesario que *B* = *C?*
c) Puesto que $A \oplus B = A \oplus C$, ¿es necesario que B = C7
Justifique las respuestas.

1.11 Para $A = \{a, b, \{a, c\}, \varnothing\}$, determine los conjuntos siguientes:
a) $A - \{a\}$
b) $A - \varnothing$
c) $A - \{\varnothing\}$
d) $A - \{a, b\}$
e) $A - \{a, c\}$
f) $A - \{\{a, b\}\}$
g) $A - \{\{a, c\}\}$
h) $\{a\} - A$
i) $\varnothing - A$
j) $\{\varnothing\} - A$
k) $\{a, c\} - A$
l) $\{\{a, c\}\} - A$
m) $\{a\} - \{A\}$

1.12 Sean *A, B y C* conjuntos arbitrarios.
a) Demostrar que *(A -B)-C* = *A* -(Bu Q
b) Demostrar que *(A-B)-C* = *(A-Q-B*
c) Demostrar que *(A* - *B)* - *C* = *{A* - Q - (B - Q

1.13 Sean los conjuntos *A, B y C.* ¿En qué condiciones son verdaderas las siguientes afirmaciones?
a) $(A - B) \cup (A - C) = A$
b) $(A - B) \cup (A - C) = \varnothing$
c) $(A - B) \cap (A - C) = \varnothing$
d) $(A - B) \oplus (A - C) = \varnothing$

1.14 Sean *A* y B dos conjuntos.
a) Si *A* - *B* = *B,* ¿qué se puede decir acerca de *A* y B?
b) Si *A* - *B* = *B* - *A,* ¿qué se puede decir acerca de *A* y B?

1.15 Señalemos con *A* el conjunto de automóviles de manufactura nacional; con *B* el conjunto de todos los automóviles importados; con C el conjunto de todos los fabricados antes de 1977; sea *D* el conjunto de todos los automóviles cuyo valor comercial actual es menor que $ 2000, y sea *E* el conjunto de todos los automóviles que son propiedad de estudiantes de la universidad. Exprese las afirmaciones siguientes en notación de teoría de conjuntos:
a) Los automóviles propiedad de estudiantes de la universidad son de manufactura nacional o importados.
b) Todos los automóviles fabricados antes de 1977 tienen un valor comercial actual menor que $ 2000.
c) Todos los automóviles importados fabricados después de 1977 tienen un valor comercial actual mayor que $ 2000.

1.16 Denotemos con *A* el conjunto de todos los estudiantes de primer año, con *B* el conjunto de todos los de segundo año, con C el conjunto de los estudiantes de la especialidad de matemáticas, con *D* el conjunto de todos los estudiantes de ciencias de la computación, con *E* el conjunto de todos los estudiantes del curso de elementos de matemáticas discretas, con *F* el de todos los que fueron al concierto de rock el lunes por la noche, y con *G* el de todos los que se acostaron tarde el lunes por la noche. Exprese las siguientes afirmaciones en notación de teoría de conjuntos:
a) Todos los estudiantes de segundo año de ciencias de la computación toman el curso de elementos de matemáticas discretas.

b) Aquellos, y sólo aquellos que están en el curso de elementos de matemáticas discretas o que fueron al concierto de rock, se acostaron tarde el lunes por la noche.

c) Ningún estudiante del curso de elementos de matemáticas discretas fue al concierto de rock el lunes por la noche (la razón obvia es la gran cantidad de problemas que dejan en el curso de elementos de matemáticas discretas).

d) El concierto de rock fue sólo para los estudiantes de primero y segundo año.

e) Todos los estudiantes de segundo año que no son de la especialidad de matemáticas ni de ciencias de la computación, fueron al concierto de rock.

1.17 Determine los conjuntos potencia de los siguientes conjuntos:

a) $\{a\}$

b) $\{\{a\}\}$

c) $\{\varnothing, \{\varnothing\}\}$

1.18 Sea $A = \{\varnothing, b\}$. Construya los siguientes conjuntos:

a) $A - \varnothing$

b) $\{\varnothing\} - A$

c) $A \cup \mathscr{P}(A)$

d) $A \cap \mathscr{P}(A)$

1.19 Sea $A = \{\varnothing\}$. Sea $B = \mathscr{P}(\mathscr{P}(A))$.

a) ¿Es $\varnothing \in B$? ¿$\varnothing \subseteq B$?

b) ¿Es $\{\varnothing\} \in B$? ¿$\varnothing \subseteq B$?

c) ¿Es $\{\{\varnothing\}\} \in B$? ¿$\{\{\varnothing\}\} \subseteq B$?

1.20 Sea $A = \{\varnothing, \{\varnothing\}\}$. Diga cuáles de las afirmaciones siguientes son verdaderas y cuáles son falsas.

a) $\varnothing \in \mathscr{P}(A)$

b) $\{\varnothing\} \subseteq \mathscr{P}(A)$

c) $\{\varnothing\} \subseteq \mathscr{P}(A)$

d) $\{\varnothing\} \subseteq A$

e) $\{\varnothing\} \in \mathscr{P}(A)$

f) $\{\varnothing\} \in A$

g) $\{\{\varnothing\}\} \subseteq \mathscr{P}(A)$

h) $\{\{\varnothing\}\} \subseteq A$

i) $\{\{\varnothing\}\} \in \mathscr{P}(A)$

j) $\{\{\varnothing\}\} \in A$

1.21 Sea $A = \{a, \{a\}\}$. Diga cuáles de las afirmaciones siguientes son verdaderas y cuáles son falsas.

a) $\varnothing \in \mathscr{P}(A)$

b) $\varnothing \subseteq \mathscr{P}(A)$

c) $\{a\} \in \mathscr{P}(A)$

d) $\{a\} \subseteq \mathscr{P}(A)$

e) $\{\{a\}\} \in \mathscr{P}(A)$

f) $\{\{a\}\} \subseteq \mathscr{P}(A)$

g) $\{a, \{a\}\} \in \mathscr{P}(A)$

h) $\{a, \{a\}\} \subseteq \mathscr{P}(A)$

i) $\{\{\{a\}\}\} \in \mathscr{P}(A)$

j) $\{\{\{a\}\}\} \subseteq \mathscr{P}(A)$

1.22 Determine si las afirmaciones siguientes son verdaderas o falsas. Explique su respuesta.

a) $A \cup \mathscr{P}(A) = \mathscr{P}(A)$

b) $A \cap \mathscr{P}(A) = A$

c) $\{A\} \cup \mathscr{P}(A) = \mathscr{P}(A)$

d) $\{A\} \cap \mathscr{P}(A) = A$

e) $A - \mathscr{P}(A) = A$

f) $\mathscr{P}(A) - \{A\} = \mathscr{P}(A)$

1.23 Sean A y B dos conjuntos arbitrarios.

a) Demuestre que $\mathscr{P}(A \cap B) = \mathscr{P}(A) \cap \mathscr{P}(B)$, o proporcione un contraejemplo.

b) Demuestre que $\mathscr{P}(A \cup B) = \mathscr{P}(A) \cup \mathscr{P}(B)$, o proporcione un contraejemplo.

1.24 Diga cuál es la cardinalidad de los conjuntos:

a) $A = \{n^7 | n$ es un entero positivo$\}$

b) $B =$ es$\{n^{109} | n$ un entero positivo$\}$

c) $A \cup B$

d) $A \cap B$

Justifique las respuestas.

1.25 Demuestre que se puede escribir en español a lo más un número infinito contable de libros (definimos un libro como una sucesión finita de palabras, dividida en oraciones, párrafos y capítulos).

1.26 *a)* Demuestre que el conjunto de todos los números racionales positivos es un conjunto infinito contable. *(Sugerencia:* Considere todos los puntos del primer cuadrante del plano cuya coordenada *x* y coordenada *y* tengan valores enteros.)

b) Demostrar que la unión de un número infinito contable de conjuntos infinitos contables es un conjunto infinito contable.

1.27 Denotemos con N el conjunto de todos los números naturales y con S el conjunto de todos los subconjuntos *finitos* de N. ¿Cuál es la cardinalidad de S? Justifique la respuesta.

1.28 *a)* Dé un ejemplo con el que demuestre que la cardinalidad de un conjunto, que es la intersección de dos conjuntos infinitos contables puede ser, a su vez, infinito contable. *b)* Dé un ejemplo con el que demuestre que la cardinalidad de un conjunto, que es la intersección de dos conjuntos infinitos contables, puede ser finita.

1.29 El señor Kantor construye una máquina que distingue los números "afortunados" de los "no afortunados". Esto es, dado un número natural, la máquina responderá: "afortunado" o "no afortunado". Dos de dichas máquinas se consideran diferentes si existe al menos un número que una máquina diga que es afortunado y que la otra diga que es no afortunado. Pruebe que hay más de un número infinito contable de dichas máquinas.

1.30 Resuelva el problema de los timbres postales del ejemplo 1.1 demostrando cómo se pueden hacer envíos de $3k$, $1k + 1$, $3k + 2$ centavos con timbres de 3 y de 5 centavos.

1.31 El señor E. V. Flores afirma que tiene una tercera parte de indígena. Cuando se le pregunta cómo es eso posible responde: "Mi padre tenía una tercera parte de indígena y mi madre tenía una tercera parte de indígena." ¿Es correcta esta demostración por inducción?

1.32 Presentamos una demostración, por inducción, de la afirmación: "Cualesquiera n bolas de billar son del mismo color."

Base de la inducción. Para n = 1, la afirmación es trivialmente verdadera.

Paso de inducción. Supongamos que tenemos $k + 1$ bolas de billar que numeramos $1, 2, \ldots, (k + 1)$. De acuerdo con la hipótesis de inducción, las bolas $1, 2, \ldots, k$ son del mismo color. Además, las bolas $2, 3, \ldots, (k + 1)$ son del mismo color. En consecuencia, las bolas $1, 2, \ldots, k, (k + 1)$ son del mismo color.

¿Dónde está el error en esta demostración?

1.33 Demuestre por inducción que para $n \geqq 0$ y $a \neq 1$

$$1 + a + a^2 + \cdots + a^n = \frac{1 - a^{n+1}}{1 - a}$$

1.34 Demuestre, por inducción, que $n^3 + 2n$ es divisible entre 3 para toda $n \geqq 1$.

1.35 Demuestre, por inducción, que $n^4 - 4n^2$ es divisible entre 3 para toda $n \geqq 2$.

1.36 Demuestre, por inducción, que 2" x 2"- 1 es divisible entre 3 para toda $n \geqq 1$.

1.37 Demuestre, por inducción, que

$$1 + 2 + 2^2 + 2^3 + \cdots + 2^n = 2^{n+1} - 1.$$

1.38 Determine la suma

$$1 + 3 + 5 + \cdots + (2n - 1)$$

a) sugiriendo una fórmula general con base en los valores de la suma para $n = 1, 2, 3, 4$ y 5; *b)* probando, por inducción, que la fórmula general es válida.

1.39 Demuestre, por inducción, que para $n \geqq 1$

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n! = (n + 1)! - 1$$

donde $n!$ es el producto $1 \cdot 2 \cdot 3 \cdots n$.

1.40 Demuestre, por inducción, que para $n \geqq 1$

$$1 \cdot 2 + 2 \cdot 3 + \cdots + n(n + 1) = \frac{n(n + 1)(n + 2)}{3}$$

1.41 Demuestre que $1^2 - 2^2 + 3^2 - 4^2 + \cdots (-1)^{n-1} n^2 = (-1)^{n-1} \dfrac{n(n + 1)}{2}$

a) Por inducción.
b) Usando el resultado del ejemplo 1.5.

1.42 Demuestre que $1^2 + 3^2 + 5^2 + \cdots + (2n - 1)^2 = \dfrac{n(2n - 1)(2n + 1)}{3}$

1.43 Demuestre que $1^3 + 2^3 + \cdots + n^3 = (1 + 2 + 3 + \cdots + n)^2$

1.44 Demuestre que $\dfrac{1^2}{1 \cdot 3} + \dfrac{2^2}{3 \cdot 5} + \cdots + \dfrac{n^2}{(2n - 1)(2n + 1)} = \dfrac{n(n + 1)}{2(2n + 1)}$

1.45 *a)* Demuestre que $\dfrac{1}{1 \cdot 2} + \dfrac{1}{2 \cdot 3} + \dfrac{1}{3 \cdot 4} + \cdots + \dfrac{1}{n(n + 1)} = \dfrac{n}{n + 1}$

b) Demuestre que $\dfrac{1}{1 \cdot 3} + \dfrac{1}{3 \cdot 5} + \cdots + \dfrac{1}{(2n - 1)(2n + 1)} = \dfrac{n}{2n + 1}$

c) Demuestre que $\dfrac{1}{1 \cdot 4} + \dfrac{1}{4 \cdot 7} + \dfrac{1}{7 \cdot 10} + \cdots + \dfrac{1}{(3n - 2)(3n + 1)} = \dfrac{n}{3n + 1}$

d) Encuentre y demuestre una fórmula general que incluya como casos particulares los resultados de *a), b) y c).*

1.46 Demuestre que

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + 3 \cdot 4 \cdot 5 + \cdots + n(n + 1)(n + 2) = \frac{n(n + 1)(n + 2)(n + 3)}{4}$$

1.47 Encuentre y demuestre, por inducción, una fórmula general que surja de la observación de que

$$1^3 = 1$$
$$2^3 = 3 + 5$$
$$3^3 = 7 + 9 + 11$$
$$4^3 = 13 + 15 + 17 + 19$$

1.48 Demuestre, por inducción, que la suma de los cubos de tres enteros consecutivos es divisible entre 9.

1.49 Demuestre que, para cualquier entero n,

$$(11)^{n+2} + (12)^{2n+1}$$

es divisible entre 133.

1.50 Se sabe que para cualquier entero positivo $n \geqq 2$

$$\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}-A>0$$

donde A es una constante. ¿Qué tan grande puede ser A?

1.51 Demuestre que para cualquier entero positivo $n > 1$

$$\frac{1}{\sqrt{1}}+\frac{1}{\sqrt{2}}+\cdots+\frac{1}{\sqrt{n}}>\sqrt{n}$$

1.52 Cuando llegan n parejas a una fiesta, el anfitrión y la anfitriona los reciben en la puerta. Después de una ronda de apretones de mano, el anfitrión pregunta a los invitados y a su esposa (la anfitriona) el número de manos que cada uno de ellos ha estrechado. Él recibió $2n + 1$ respuestas diferentes. Puesto que nadie estrechó la mano de su esposo o esposa, ¿cuántas manos estrechó la anfitriona? Demuestre el resultado por inducción.

1.53 *a)* Sea S el conjunto de números naturales tales que: 1. El número natural n_0 está en S. 2. Si los números naturalesn_0, $n_0 +1$, $n_0 + 2$,..., k están en S, entonces el número natural k+\ también está en S.

Demuestre que S es el conjunto de todos los números naturales mayores o iguales que n_Q. *Sugerencia:* suponga que n, es el menor número natural que no está en S. *b)* Use el resultado de la parte *a)* para demostrar que, en efecto, el principio de la inducción matemática fuerte es válido.

1.54 Entre los enteros del 1 al 300, ¿cuántos no son divisibles entre 3, ni entre 5, ni entre 7? ¿Cuántos son divisibles entre 3 pero no entre 5 ni entre 7?

1.55 Se van a distribuir al azar N Juguetes entre N niños. Hay una manera interesante en que los niños elijan los juguetes de modo que ningún par de ellos escoja el mismo juguete. Se traza una gráfica como la mostrada en la figura 1P.1*a*) donde hay N líneas verticales y un número arbitrario de segmentos horizontales al azar entre las líneas verticales adyacentes con la condición de que ningún par de segmentos horizontales se encuentre en el mismo punto. Se asignan los TV juguetes a la parte inferior de las líneas verticales y cada niño escoge como punto de partida la parte superior de una línea vertical. Desde este punto, el nifio trazará una trayectoria hacia abajo. Sin embargo,

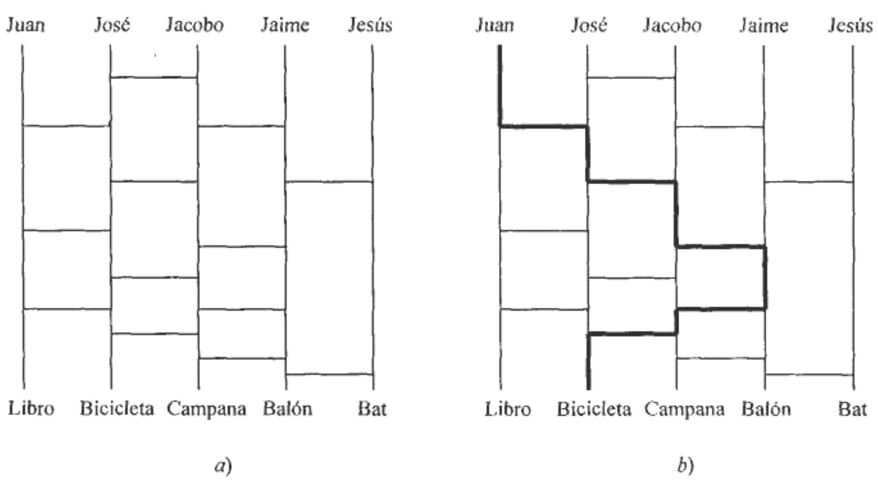


Figura 1P.1

cada vez que el niño encuentre un segmento horizontal deberá continuar horizontalmente y doblará hacia abajo cuando encuentre la línea vertical adyacente. Por ejemplo, la figura IR 16) muestra la trayectoria que sigue Juan. Se afirma que sin importar cuántos segmentos horizontales se tracen, de cualquier manera ningún par de niños llegará al mismo juguete. Demuestre esta afirmación por inducción sobre el número de segmentos horizontales trazados.

1.56 Se realizó una encuesta entre 1000 personas. De éstas, 595 son demócratas, 595 usan anteojos y a 550 les gustan los helados; 395 de ellas son demócratas que usan anteojos, 350 son demócratas a quienes les gustan los helados y 400 usan anteojos y les gustan los helados; 250 de ellas son demócratas que usan anteojos y a quienes les gustan los helados. ¿Cuántos no son demócratas, no usan anteojos y no gustan de los helados? ¿Cuántos son demócratas que no usan anteojos y a quienes no les gustan los helados?

1.57 Se sabe que en la universidad el 60% de los profesores juega tenis, el 50% juega *bridge,* el 70% corre, el 20% juega tenis y *bridge,* el 30% juega tenis y corre, y el 40% juega *bridge* y corre. Si alguien afirma que el 20% de los profesores corre y juega *bridge y* tenis, ¿lo creería? ¿Por qué?
Los 60 000 aficionados que asistieron al partido de fútbol compraron toda la parafernalia *ad hoc* para sus autos. En total se vendieron 20 000 letreros para defensas, 36 000 calcomanías para ventanas y 12 000 llaveros. Sabemos que 52 000 aficionados compraron al menos un artículo y ninguno compró más de un artículo dado. También, que 6000 compraron calcomanías y llaveros, 9 000 compraron calcomanías y letreros y que 5000 compraron llaveros y letreros.

a) ¿Cuántos aficionados compraron los tres artículos?
b) ¿Cuántos compraron exactamente un artículo?
c) Alguien puso en duda la precisión del número total de compradores: 52 000 (ya que se verificaron las otras cifras). Esta persona afirmó que el número total de compradores es de 60 000 o 44 000. ¿Cómo desmiente esa afirmación?

1.59 De un total de 130 estudiantes, 60 usan sombrero, 51 usan bufanda y 30 usan ambas prendas: sombrero y bufanda. De los 54 estudiantes que usan suéter, 26 usan sombrero, 21 usan bufanda y 12 usan tanto sombrero como bufanda. Quien no usa ni sombrero ni bufanda, usa guantes.

a) ¿Cuántos estudiantes usan guantes?
b) ¿Cuántos estudiantes que no usan suéter usan sombrero pero no bufanda?
c) ¿Cuántos estudiantes que no usan suéter no usan sombrero ni bufanda?

1.60 De 100 estudiantes, 32 estudian matemáticas, 20 estudian física, 45 estudian biología, 15 estudian matemáticas y biología, 7 estudian matemáticas y física, 10 estudian física y biología, y 30 no estudian ninguna de estas tres materias.

a) Encuentre el número de estudiantes que estudian las tres materias.
b) Encuentre el número de estudiantes que estudian exactamente una de las tres materias.

1.61 En una reunión de DAR (*Daughters of the American Revolution:* Hijas de la Revolución Americana) de 30 mujeres, 17 son descendientes de George Washington, 16 son descendientes de John Adams y 5 no son descendientes de Washington ni de Adams. ¿Cuántas de las 30 mujeres son descendientes tanto de Washington como de Adams?

1.62 Setenta y cinco niños fueron a un parque de diversiones donde subieron a la rueda de la fortuna, la montaña rusa y al trenecito. Se sabe que 20 de ellos subieron a los tres juegos y que 55 subieron al menos a dos de esos tres juegos. Cada juego cuesta \$ 0.50 y el costo del total fue de \$ 70. Determine el número de niños que no subió a ninguno de los juegos.

1.63 *a)* De 50 estudiantes de una clase, 26 obtuvieron A en el primer examen y 21 obtuvieron A en el segundo. Si 17 estudiantes no obtuvieron A en ninguno de los dos exámenes, ¿cuántos estudiantes obtuvieron A en ambos exámenes?

b) Si el número de estudiantes que obtuvo A en el primer examen es igual al del segundo examen, si el número total de los estudiantes que obtuvieron A en exactamente un examen es 40 y si 4 estudiantes no obtuvieron A en ningún examen, encuentre el número de estudiantes que obtuvieron A sólo en el primer examen, de los que obtuvieron A sólo en el segundo examen, y de los que obtuvieron A en ambos exámenes.

1.64 Una posible manera de definir la diferencia simétrica de dos multiconjuntos *P* y *Q*, denotada por $P \oplus Q$, es establecer que la multiplicidad de un elemento de $P \oplus Q$ es igual al valor absoluto de la diferencia entre las multiplicidades del elemento en *P y* en *Q*. ¿Cuál es una posible inconsistencia con dicha definición?
Sugerencia: Considere los multiconjuntos $(P \oplus Q) \oplus R$ y $P \oplus (Q \oplus R)$.

1.65 Para describir los diversos restaurantes en la ciudad, denotemos con *p* la afirmación "la comida es buena", con *q* la afirmación "el servicio es bueno" y con r la afirmación "es de tres estrellas". Escriba las afirmaciones siguientes en forma simbólica.

a) La comida es buena o el servicio es bueno, o ambas cosas.
b) La comida es buena o el servicio es bueno, pero no ambas cosas.
c) La comida es buena pero el servicio no.
d) No sucede que tanto la comida sea buena como que el restaurante sea de tres estrellas.
e) Si tanto la comida como el servicio son buenos, entonces el restaurante es de tres estrellas.
f) No es cierto que ser de tres estrellas siempre signifique buena comida y buen servicio.

1.66 Denotemos con *p* la afirmación "el material es interesante", con *q* la afirmación "los ejercicios son difíciles" y con *r* la afirmación "el curso es agradable". Escriba las afirmaciones siguientes en forma simbólica:

a) El material es interesante y los ejercicios son difíciles.
b) El material no es interesante, los ejercicios no son difíciles y el curso no es agradable.
c) Si el material no es interesante y los ejercicios no son difíciles, entonces el curso no es agradable.
d) Que el material sea interesante significa que los ejercicios son difíciles, y viceversa.
e) O el material es interesante o los ejercicios no son difíciles, pero no ambas cosas.

1.67 Escriba las afirmaciones siguientes en forma simbólica:

a) El sol brilla y la humedad no es alta.
b) Si termino mi tarea antes de la cena y no llueve, entonces iré al partido de fútbol.
c) Si no me ves mañana significa que habré ido a la playa.
d) Si el costo de las utilidades crece o se niega la requisición de fondos adicionales, entonces compraremos una nueva computadora si, y sólo si, podemos mostrar que los recursos de cómputo son, en efecto, insuficientes.

1.68 Denotemos con *p* la afirmación "el clima es agradable" y con *q* la afirmación "vamos de día de campo". Traduzca lo siguiente al español y simplifique si es posible:

a) $p \wedge \overline{q}$
b) $p \leftrightarrow q$
c) $\overline{q} \rightarrow \overline{p}$
d) $\overline{(\overline{p} \vee q)} \vee (p \wedge \overline{q})$

1.69 *d)* Escriba una afirmación compuesta que sea verdadera cuando exactamente dos de tres afirmaciones *p*, *q* y *r* sean verdaderas.
b) Escriba una afirmación compuesta que sea verdadera cuando ninguna, o una, o dos de las tres afirmaciones *p*, *q* y *r* sean verdaderas.

1.70 Construya las tablas de verdad de las siguientes afirmaciones:

a) $p \rightarrow p$
b) $(p \rightarrow p) \vee (p \rightarrow \overline{p})$
c) $(p \rightarrow p) \rightarrow (p \rightarrow \overline{p})$
d) $(p \vee \overline{q}) \vee \overline{p}$
e) $(p \vee \overline{q}) \rightarrow \overline{p}$
f) $p \leftrightarrow (\overline{p} \vee \overline{q})$

g) $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$
h) $(\overline{q} \to \overline{p}) \to (p \to q)$

1.71 *a)* Dado que el valor de $p \to q$ es falso, determine el valor de $(\overline{p} \vee \overline{q}) \to q$.
b) Dado que el valor de $p \to q$ es verdadero, ¿puede determinar el valor de $\overline{p} \vee (p \leftrightarrow q)$?

1.72 Considere los siguientes anuncios sobre un juego:
a) Hay tres afirmaciones en este anuncio.
b) Dos de ellas no son verdaderas.
c) El promedio de crecimiento del IQ de las personas que aprenden este juego es de más de 20 puntos.

¿Es verdadera la afirmación *c*)?

1.73 Cierto país está habitado por personas que siempre dicen la verdad o que siempre mienten, y que responderán preguntas sólo con "sí" o "no". Un turista llega a una bifurcación en el camino, una de cuyas ramas conduce a la capital y la otra no. No hay un letrero que diga qué camino seguir, pero hay un nativo, el señor Z, parado en la bifurcación. ¿Qué única pregunta deberá hacerle el turista para determinar qué camino seguir?

Sugerencia: Sea *p* "el señor Z siempre dice la verdad" y *q* "el camino de la izquierda conduce a la capital". Formule una proposición *A* que incluya a *p* y a *q*, tal que la respuesta del señor Z a la pregunta *"¿A* es verdadera?" sea "sí" cuando, y sólo cuando, *q* sea verdadera.

1.74 Antonio, Miguel y Juan pertenecen al Club Alpino. Cada miembro del club esquía, escala o ambas cosas. A ningún escalador le gusta la lluvia y a todos los esquiadores les gusta la nieve. A Miguel le disgusta lo que a Antonio le gusta, y le gusta lo que a Antonio le disgusta. A Antonio le gusta la lluvia y la nieve. ¿Hay algún miembro del Club Alpino que sea escalador pero no esquiador?

CAPÍTULO

DOS

Conmutabilidad y lenguajes formales

2.1 INTRODUCCIÓN

En este capítulo examinaremos una cuestión que en apariencia es simple, a saber, "¿Cómo especificamos los elementos de un conjunto?" En la sección 1.1 sugerimos dos maneras sencillas de especificar los elementos de un conjunto: listar exhaustivamente todos los elementos del conjunto (siempre que, por supuesto, el conjunto sea finito) y especificar las propiedades que caracterizan de manera exclusiva a los elementos del conjunto. Por ejemplo,

$$\{9,16,25,36,49\}$$

y

$$\{x \mid x \text{ es un cuadrado perfecto, } 5 \le x \le 50\}$$

son dos maneras de describir el mismo conjunto. Hasta ahora hemos concedido que esto es correcto. Por un lado, no hemos analizado con detalle si puede haber alguna falla cuando especificamos los elementos de un conjunto de estas dos maneras. Por el otro, no hemos buscado otras maneras posibles que también puedan ser efectivas y útiles. Sucede que algunos de los temas que estudiaremos en este capítulo no sólo son muy interesantes sino que también tienen gran importancia en ciencias teóricas de la computación.

2.2 LA PARADOJA DE RUSSELL Y NO COMPUTABILIDAD

Cuando queremos especificar los elementos de un conjunto que contiene sólo unos cuantos elementos, la manera más directa y obvia es listar exhaustivamente todos los elementos del conjunto. Sin embargo, cuando un conjunto contiene un gran número o un número infinito de elementos, resulta impráctico o imposible listar todos los elementos del conjunto. Por ejemplo, podemos tener

$$P = \{x \mid x \text{ es un estudiante de secundaria en Caracas}\}$$

donde P es un conjunto finito con un gran número de elementos. Podemos tener

$$Q = \{x \mid x \text{ es un cuadrado perfecto}\}$$

donde Q es un conjunto infinito contable de enteros. También podemos tener

$$R = \{x \mid \{a, b\} \subseteq x\}$$

R es un conjunto de conjuntos tal que todo elemento de R contiene al conjunto $\{a, b\}$ como subconjunto.

Queremos mostrar que hay una posible falla cuando especificamos los elementos de un conjunto al especificar las propiedades que caracterizan de manera exclusiva a estos elementos. Considere el conjunto

$$S = \{x \mid x \notin x\}$$

Parece que hemos seguido una "receta" y hemos definido un conjunto S tal que un conjunto x es elemento de S si x «? x. Así, por ejemplo, $\{a, b\}$ es un elemento de S, pues $\{a, b\} \notin \{a, b\}$. También $\{\{a\}\}$ es un elemento de S, pues $\{\{a\}\} \notin \{\{a\}\}$. Sin embargo, supongamos que alguien quiere saber si S es un elemento de S. En otras palabras, esa persona quiere saber si $S \in S$. Siguiendo la especificación, decimos: para que S sea un elemento de S debe darse el caso de que $S \notin S$, lo cual es una afirmación autocontradictoria. Hagamos lo contrario y supongamos que S no es un elemento de S; esto es, $S \notin S$. Entonces, de acuerdo con la especificación, S deberá ser un elemento de S. Esto es, si $S \notin S$ entonces de $S \in S$, nuevo una afirmación autocontradictoria. Aclaramos que lo dicho con anterioridad no es un juego de palabras y, de ninguna manera, un intento de confundir al lector con sintaxis difícil y complicada. Más bien, al contrario de lo que podría pensarse intuitivamente, no siempre sucede que podemos especificar con precisión los elementos de un conjunto especificando las propiedades que tienen sus elementos. Dicha observación la efectuó por primera vez B. Russell en 1911, y se conoce como *paradoja de Russell.*

Para que el lector tenga una apreciación más profunda del argumento anterior, proporcionamos otros ejemplos:

Ejemplo 2.1

Hay un barbero en un pueblo. Él afeitará a todos los que no se afeiten a sí mismos. Parece ser que tenemos una descripción precisa de cierto barbero del pueblo. Sin embargo, supongamos que surge la pregunta de si el barbero se afeita a sí mismo. Si el barbero se afeita a sí mismo entonces él no deberá afeitarse (pues sólo afeita a quienes no se afeitan a sí mismos). Por otro lado, si el barbero no se afeita a sí mismo, entonces él deberá afeitarse (pues él afeita a los que no se afeitan a sí mismos). En ambos casos tenemos una afirmación autocontradictoria. De nuevo, la falla es la manera en que especificamos al barbero. En consecuencia, concluimos que no puede existir dicho barbero.

❐

Ejemplo 2.2

Definimos una propiedad que un adjetivo puede tener o no. Se dice que un adjetivo es *heterológico* si no posee la propiedad que describe. Por ejemplo, el adjetivo "monosilábico" es heterológico (pues tiene más de una sílaba), mientras que el adjetivo "polisilábico" no es heterológico (pues tiene más de una sílaba). También el adjetivo "largo" es heterológico si acordamos que una palabra de seis o más letras es una palabra larga. Análogamente, el adjetivo "corto" es heterológico. Preguntamos ahora "¿El adjetivo 'heterológico' es heterológico?" Si el adjetivo "heterológico" es heterológico, entonces no es heterológico. Por otro lado, si el adjetivo "heterológico" no es heterológico, entonces es heterológico. En consecuencia, concluimos que heterológico no es una propiedad bien definida.

Ejemplo 2.3

Hay dos artesanos, Bellini y Cellini, en Florencia. Sobre cualquier cosa que hace, Bellini coloca una inscripción verdadera. Por otro lado, Cellini coloca una inscripción falsa sobre lo que hace. Si ellos son los únicos artesanos en los alrededores, ¿qué diría si alguien asegura que vio el siguiente letrero?

Este letrero
lo hizo
Cellini

Ahora damos una ilustración final del argumento estudiado. Aunque todos estamos impresionados por la potencia y versatilidad de la computadora, vamos a mostrar que hay tareas que ninguna computadora puede efectuar. En efecto, éste es uno de los conceptos más importantes de las ciencias teóricas de la computación. (De nuevo, pedimos al lector que antes de continuar piense en la pregunta: "¿cómo probamos que ninguna computadora es capaz de realizar cierta tarea?" Para probar que una computadora es capaz de realizar cierta tarea podemos, simplemente, mostrar la manera como se puede realizar la tarea. Por otro lado, cuando decimos que cierta tarea está más allá de la capacidad de todas las computadoras, ¿cómo sabemos que ello no se debe a nuestra ineptitud como usuarios de la computadora? Quizás alguien más pueda hallar una manera de usar la computadora para que realice la tarea. ¿Cómo sabemos con certeza que ello no se debe al hecho de que nuestra computadora no es suficientemente "poderosa" para realizar la tarea? Quizás alguien más tenga una computadora que sea lo suficientemente poderosa para efectuar la tarea.) Una de las pesadillas de cualquier estudiante de ciencias de la computación es que un programa entre en un "lazo infinito" y nunca pare. Así, los estudiantes con frecuencia buscan ayuda de algún asesor amable a quien piden que vea el programa y los datos que usa, para determinar si el programa parará alguna vez. En lugar de imponer tan desagradable actividad al asesor, nos preguntamos sobre la posibilidad de escribir un programa que examine cualquier programa de los estudiantes, con los datos que usa, y determine si el programa, trabajando con los datos obtenidos, parará alguna vez. *Supongamos* que alguien ha escrito dicho programa, al cual

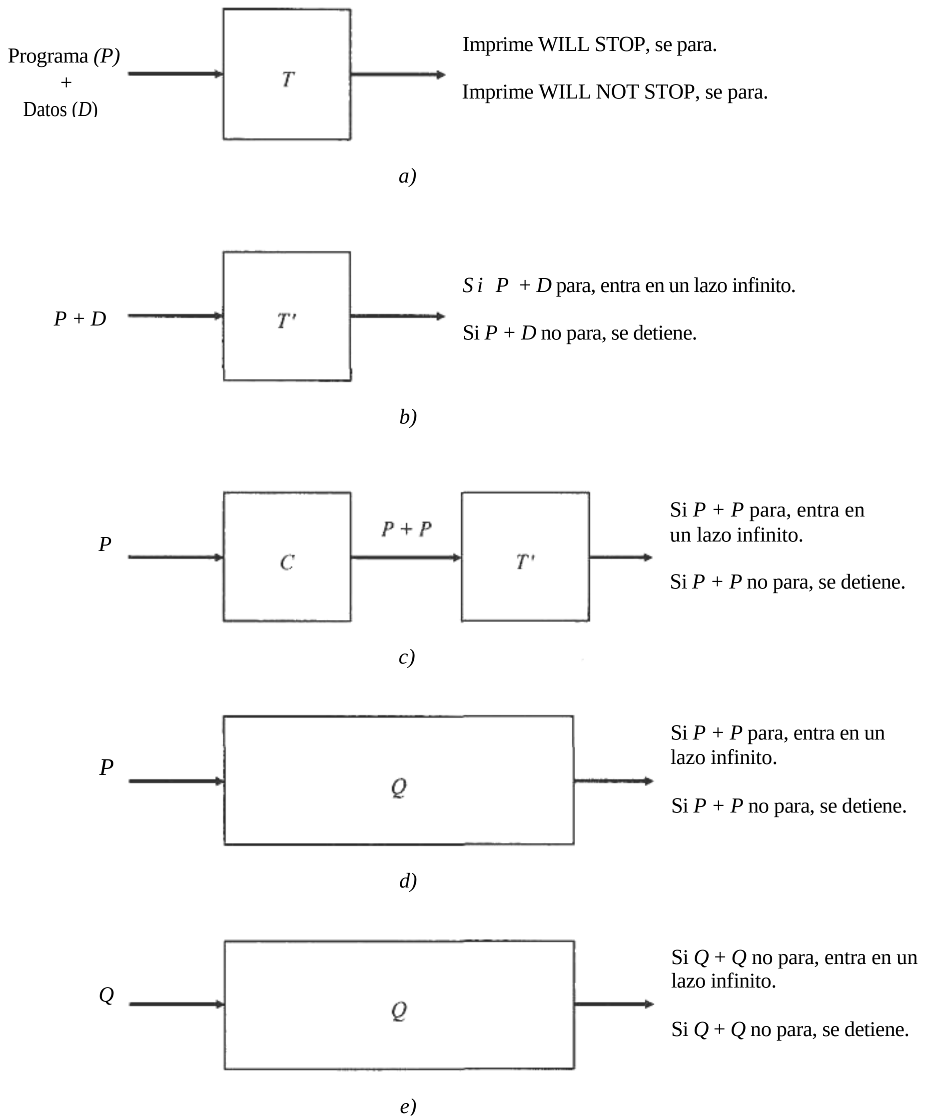


Figura 2.1

nos referiremos como *T*. El comportamiento del programa *T´* se puede describir en la figura 2.1*a*, a saber, el programa *T* examinará un programa *P y* los datos *D* con que trabaja *P*, efectuará el diagnóstico correcto, imprimirá el mensaje correspondiente y después parará. Dado el programa *T*, podemos hacer una ligera modificación para obtener un programa *T´*. La única diferencia entre los programas *T y T´* es que después de diagnosticar que un programa que trabaja con un conjunto dado de datos parará, el programa *T´* mismo entrará en un lazo infinito. La figura 2.1*b* describe el comportamiento del programa *T´*. (Dicha modificación es sencilla, pues todos sabemos cómo añadir un lazo infinito a un programa.)

Un programa está compuesto de cadenas de letras y dígitos, por tanto, lo mismo sucede con los datos; es posible que un programa se use a sí mismo como el conjunto de datos con los que trabaja. (El que los resultados tengan sentido o no, es otro asunto, del cual no nos ocupamos aquí.) Así, al construir un pequeño programa *C* que copie un programa para que el programa lo use como datos, tenemos lo que se muestra en la figura 2.1*c*. Coloquemos juntos los programas *C* y *T´* y nos referiremos a ellos como el programa *Q;* esto se muestra en la figura 2.1c/. Podemos resumir el comportamiento del programa *Q* así:

1. El programa *Q* acepta al programa *P* como entrada.
2. Si el programa *P*, cuando usa al programa *P* como dato de entrada, en algún momento va a parar, entonces el programa *Q* entrará en un lazo infinito.
3. Si el programa *P,* cuando usa al programa *P* como dato de entrada, va a entrar en un lazo infinito, entonces el programa *Q* parará.

Como el mismo programa *Q* es un programa, podemos hacer que el programa *Q* sea la entrada del programa *Q.* Así, tenemos la situación descrita en la figura 2.1 *e.* Ahora, de acuerdo con las afirmaciones 2 y 3 anteriores, tenemos:

2'. Si el programa *Q,* cuando usa al programa *Q* como dato de entrada, en algún momento va a parar, entonces el programa *Q* entrará en un lazo infinito.

3'. Si el programa *Q,* cuando usa al programa *Q* como dato de entrada, va a entrar en un lazo infinito, entonces el programa *Q* parará.

Es evidente que las afirmaciones 2' y 3' son autocontradictorias. De nuevo, lo que conduce a dicha afirmación contradictoria es la suposición de que existe ese programa *T.* En consecuencia, podemos concluir que no es posible escribir un programa que examine cualquier programa y conjunto de datos dado y determine si el programa en algún momento va a parar.

Debemos concluir esta sección restableciendo la confianza del lector. Resulta (aunque el análisis detallado del tema rebasa el ámbito de este libro) que no habrá confusión al especificar un conjunto mediante propiedades que caracterizan a los elementos del conjunto, en tanto nos restrinjamos a cierto *universo* de elementos. Por ejemplo, denotemos con *A* el conjunto de todos los estudiantes. Entonces es claro que los siguientes conjuntos están bien definidos:

$$F = \{x \mid x \in A, x \text{ es de segundo año}\}$$

$$G = \{x \mid x \in \mathscr{P}(A), x \text{ es un conjunto de estudiantes de un grupo real}\}$$

En el primer ejemplo, *A* es el universo. En el segundo ejemplo $\mathscr{P}(A)$ es el universo. En este momento un observador perspicaz señalaría que la paradoja de Russell surgirá de nuevo si usamos como universo un conjunto que contenga todos los conjuntos del mundo. Sin embargo, se puede demostrar que no existe un conjunto que contenga todos los conjuntos del mundo (¿dicho conjunto se contiene a sí mismo como elemento?). En efecto, la paradoja de Russell es equivalente a la hipótesis de la existencia de dicho conjunto.

2.3 CONJUNTOS ORDENADOS

Como lo mencionamos en el capítulo 1, un conjunto es una colección *no ordenada* de objetos. En esta sección introduciremos el concepto de conjuntos ordenados de objetos. Para ello, primero definimos el concepto de par ordenado de objetos. Un *par ordenado* de objetos es un par de objetos arreglado en un orden fijo. Así, es posible referirse a estos dos objetos como al primer objeto y al segundo objeto del par ordenado. Usamos la notación *(a, b)* para denotar el par ordenado en el cual el primer objeto (componente) es a y el segundo objeto (componente) es *b.* Un par ordenado difiere de dos maneras de un conjunto de dos objetos: primero, el orden de los dos objetos en un par ordenado es importante. Así, *(a, b)* y *(b, a)* son dos pares ordenados diferentes. Segundo, los dos objetos en un par ordenado no tienen que ser distintos. Así, *(a, a)* es un par ordenado bien definido. A menudo encontramos el concepto de pares ordenados de objetos; por ejemplo, entre todos los jugadores en un torneo de tenis, un par ordenado *(a, b)* puede denotar al campeón y al retador del torneo. Así, el par ordenado *(a, b)* no significa lo mismo que el par ordenado *(b, a).* Como otro ejemplo, entre todos los estudiantes de una clase, un par ordenado *(a, b)* puede denotar a los estudiantes que obtuvieron la calificación más alta en dos exámenes. Así, el par *(a, a)* significa que el estudiante *a* obtuvo la más alta calificación en ambos exámenes.

El concepto de pares ordenados se puede extender de inmediato. Aunque es intuitivamente obvio que podemos querer definir una terna ordenada de objetos *(a, b, c),* como una terna de objetos en la cual el primero es *a,* el segundo es *b* y el tercero es c, en vez de ello definimos de modo formal una *terna ordenada* como un par ordenado *((a, b), c)*, donde la primera componente del par ordenado es, a su vez, un par ordenado.† De manera similar, definimos una *cuarteta ordenada* como un par ordenado *(((a, b), c), d),* donde la primera componente del par ordenado es una terna ordenada. También definimos una *n-ada ordenada* como un par ordenado donde la primera componente es una (« - l)-ada ordenada.

Una *n*-ada ordenada es, en efecto, un "conjunto ordenado" con *n* elementos. En específico el 1^o, 2°,…,(*n*- l)-avo elementos del conjunto ordenado son los *n* - 1 elementos de la primera componente de la *n*-ada ordenada que, repetimos, es una (*n* -l)-ada ordenada, y el H-ésimo elemento del conjunto ordenado es la segunda componente de la *n*-ada ordenada. Así, para una *n*-ada ordenada, nos podemos referir al primero, segundo y *n*-ésimo elemento (componente) de la *n*-ada ordenada. En la práctica, con frecuencia relajamos la notación y usamos *(a, b, c)* o *abc* para representar la terna ordenada *((a, b), c);* usamos *(a, b, c, d)* o *abcd* para representar la cuarteta ordenada *(((a, b), c), d),* y así sucesivamente.

2.4 LENGUAJES

Denotemos con $A = \{a, b, c, d, \ldots, x, y, z\}$ las 27 letras del alfabeto español. Naturalmente, una palabra de *n* letras es una *n*-ada ordenada de letras del alfabeto. En efecto, la quinteta

† La razón de usar dicha definición formal es que no se necesita introducir un *nuevo* concepto.

ordenada (((($o, r), d), e), n$) corresponde a la palabra *orden.* (Ésta es la razón exacta por la cual introdujimos la notación más sencilla para las n-adas ordenadas en la sección 2.3.) En el contexto de lenguajes a menudo usamos los términos *sucesiones, cadenas* o *enunciados* (de letras) de manera intercambiable con el término n-adas ordenadas (de letras). Usamos la notación A^n o $\{a, b,c,d,\ldots ,x, y, z\}^n$ para denotar el *conjunto* de todas las sucesiones de n letras del conjunto A o del conjunto $\{a, b, c, d, \ldots, x, y, z\}$. También usamos la notación $A^{\star}$ o $\{a, b, c, d, \ldots, x, y, z\}^{\star}$ para denotar el conjunto de todas las sucesiones de una letra, de dos letras y de n letras.† Así, por ejemplo, el conjunto de todos los nombres de un directorio telefónico es un subconjunto de $A^{\star}$. Además, el conjunto de todos los nombres de 5 letras del directorio es un subconjunto de A^5. Como otro ejemplo, sea B un conjunto formado por las mayúsculas y minúsculas de las 27 letras del alfabeto español, así como por los 7 signos de puntuación: el punto, la coma, los dos puntos, el punto y coma, los signos de admiración y de interrogación, y el guión (que denota un espacio en blanco). Esto es,

$$B = \{ a, b, \ldots, y, z, A, B, \ldots, Y, Z, ., ,, :, ;, ¡, ?, \_\}$$

Es claro que un enunciado en español como ¿Dónde está Juan? es una sucesión de 5*. De manera análoga, una afirmación en cualquier lenguaje de programación es una sucesión de $C^{\star}$, donde C es el conjunto de caracteres de ese lenguaje de programación particular, como

$$C= \{A, B,\ldots, Y, Z, 0, 1, 2,\ldots, 8, 9, +, -, *, /,;,., =\}$$

Definimos de modo formal el concepto de lenguaje: sea A un conjunto finito, que es el *alfabeto* del lenguaje. Un *lenguaje* (sobre el alfabeto A) es un subconjunto del conjunto $A^{\star}$.‡ Así, por ejemplo, sea $A = \{a, b, c\}$. Todos los conjuntos siguientes son lenguajes sobre el alfabeto A.

$$L_1 = \{a, aa, ab, ac, abc, cab\}$$
$$L_2 = \{aba, aabaa\}$$
$$L_3 = \{\ \ \}$$
$$L_4 = \{a^icb^i \mid i \geq 1\}$$

En la especificación de L_4 usamos a^i para denotar una sucesión de i letras a. Así, a^icb^i quiere decir una sucesión de i letras a seguida de c, seguida de i letras b.

Los lenguajes se definen como conjuntos de cadenas; por tanto, se pueden aplicar todas las operaciones de conjuntos a los lenguajes. Por ejemplo, sean L_1 y L_2 dos lenguajes, $L_1 \cup L_2$ también es un lenguaje, que contiene los enunciados que están en L, o en L_2. Así, si L, es el lenguaje español y L_2 es el lenguaje francés, $L_1 \cup L_2$ será el conjunto de todos los enunciados que alguien que hable español y francés puede reconocer. De manera similar, $L_1 \cap L_2$ es un

† En otras publicaciones a menudo se usa la notación A^+ o $\{a, b, c, d, \ldots, x, y, z\}^+$. Usamos $\star$ para evitar una posible confusión con la notación empleada para el conjunto sucesor de un conjunto, estudiada en el capítulo 1.

‡ En el presente estudio no usamos el concepto de "sucesión nula", esto es, una sucesión que no contiene letras. En general, un lenguaje puede contener la sucesión nula como una de sus sucesiones.

lenguaje que contiene todos los enunciados que están en ambos, L_1 y L_2 Así, sea L_1 el lenguaje de programación FORTRAN, y sea L_2 el lenguaje de programación Pascal. Entonces $L_1 \cap L_2$ será el conjunto de todos los enunciados que son válidos tanto en FORTRAN como en Pascal. También, como otros ejemplos, observe que

$$\{a^i b^j \mid i > j \geqq 1\} \cup \{a^i b^j \mid 1 \leqq i < j\} = \{a^i b^j \mid i \neq j, i, j \geqq 1\}$$

$$\{a^i b^i c^j \mid i, j \geqq 1\} \cap \{a^i b^j c^j \mid i, j \geqq 1\} = \{a^i b^i c^i \mid i \geqq 1\}$$

$$\{a^i b^j \mid i \geqq j \geqq 1\} \oplus \{a^i b^j \mid 1 \leqq i \leqq j\} = \{a^i b^j \mid i \neq j, i, j \geqq 1\}$$

$$\{a^i b^j \mid i > j \geqq 1\} - \{a^i b^j \mid i \geqq 1\} = \{a^i b^j \mid i \neq j, i, j \geqq 1\}$$

2.5 GRAMÁTICAS DE ESTRUCTURAS DE FRASES

Como un lenguaje es un conjunto de cadenas, el problema de especificar un lenguaje ya no es un problema nuevo. Sin embargo, las dos maneras de especificar los elementos de un conjunto, propuestas en el capítulo 1, no son adecuadas para el caso de los lenguajes. Debido a que la mayoría de los lenguajes, si no es que todos, tienen un número infinito de cadenas, no es posible efectuar un listado exhaustivo de todas las cadenas. Incluso para cualquier lenguaje no trivial, es muy complicado describir las propiedades que caractericen de manera única a todas las cadenas del lenguaje. Más aún, para muchas aplicaciones estamos interesados principalmente en los siguientes problemas:

1. Dada la especificación de un lenguaje, generar automáticamente una o más cadenas del lenguaje.
2. Dada la especificación de un lenguaje, determinar si una cadena dada está en el lenguaje.

Por tanto, conviene tener maneras de describir lenguajes que faciliten la solución de estos problemas. Para ello introducimos el concepto de especificar un lenguaje mediante una gramática, idea originada del estudio de los lenguajes naturales. En particular, estudiaremos una clase de gramáticas, conocidas como *gramáticas de estructuras de frases.*

Presentamos un ejemplo para motivar las definiciones formales que estudiaremos posteriormente. Supongamos que nos limitamos a un subconjunto muy restringido de los enunciados en español. Podemos comenzar preguntando lo que es un enunciado en el lenguaje. Supongamos que se puede tener un enunciado en una de dos formas:

1. Un **enunciado** es una **frase nominal** seguida de una **frase-de-verbo-transitivo** y de otra **frase nominal.**
2. **Un enunciado** es una **frase nominal** seguida de una **frase-de-verbo-intransitivo.**

En seguida nos preguntamos qué son frase nominal, frase-de-verbo-transitivo y frase-de-verbo-intransitivo. Podemos tener:

3. Una **frase nominal** es un **artículo** seguido de un **sustantivo.**
4. Una **frase nominal es** un **sustantivo.**

También,

5. Una **frase-de-verbo-transitivo** es **un verbo transitivo.**

También,

6. Una **frase-de-verbo-intransitivo** es **un verbo intransitivo** seguido de un **adverbio.**
7. Una **frase-de-verbo-intransitivo** es **un verbo intransitivo.**

Ahora debemos especificar lo que es un artículo, un sustantivo, un verbo transitivo, un verbo intransitivo y un adverbio. Podemos tener:

8. Un **artículo** es ***un.***
9. **Un artículo es** *el.*

También,

10. Un **sustantivo** es perro.

11. Un **sustantivo** es *gato.*

También,

12. Un **verbo transitivo** es *persigue.*

13. Un **verbo transitivo** es *halla.*

También,

14. Un **verbo intransitivo** es *corre.*

También,

15. Un **adverbio** es *lentamente.*
16. Un **adverbio** es *rápidamente.*

Todo esto genera la siguiente notación:

enunciado → **frase nominal frase-de-verbo-transitivo frase nominal**
enunciado → **frase nominal frase-de-verbo-intransitivo**
frase nominal → **artículo sustantivo**
frase nominal → **sustantivo**
frase-de-verbo-transitivo → **verbo transitivo**
frase-de-verbo-intransitivo → **verbo intransitivo adverbio**
frase-de-verbo-intransitivo → **verbo intransitivo**

$$
\begin{aligned}
\textbf{artículo} &\rightarrow \textit{un} \qquad \textbf{(2.1)}\\
\textbf{artículo} &\rightarrow \textit{el}\\
\textbf{sustantivo} &\rightarrow \textit{perro}\\
\textbf{sustantivo} &\rightarrow \textit{gato}\\
\textbf{verbo transitivo} &\rightarrow \textit{persigue}\\
\textbf{verbo transitivo} &\rightarrow \textit{halla}\\
\textbf{verbo intransitivo} &\rightarrow \textit{corre}\\
\textbf{adverbio} &\rightarrow \textit{lentamente}\\
\textbf{adverbio} &\rightarrow \textit{rápidamente}
\end{aligned}
$$

El significado de la flecha en las líneas anteriores ha de ser obvio. Indica la posibilidad de transformar lo que está en su lado izquierdo en lo que está en su lado derecho. Además, el lector puede ver inmediatamente que algunos enunciados del lenguaje son:

el perro halla un gato
el perro persigue un gato
el gato corre lentamente

Explicaremos muchos detalles adicionales después de presentar algunas notaciones generales.

Como se definió anteriormente, un lenguaje es un subconjunto de las cadenas de $A^{\star}$. Se puede usar una *gramática de estructuras de frases* para especificar un lenguaje. Consta de cuatro elementos:

1. Un conjunto de *terminales T*.
2. Un conjunto de *no terminales N*.
3. Un conjunto de *producciones P*.
4. De entre las no terminales en *N*, hay una no terminal especial que se conoce como *símbolo inicial.*

Explicamos en seguida lo que esto significa:

1. Las terminales en *T* son símbolos usados para formar enunciados en el lenguaje. Para el ejemplo anterior, el conjunto *{un, el, perro, gato, persigue, halla, corre, lentamente, rápidamente)* es el conjunto de terminales.
2. Las no terminales en *N* son los símbolos intermedios usados para describir la estructura de los enunciados. En el ejemplo anterior, el conjunto **{enunciado, frase nominal, sustantivo, artículo, frase-de-verbo-transitivo, verbo transitivo, frase-de-verbo-intransitivo, verbo intransitivo, adverbio}** es el conjunto de no terminales.

3. Las producciones son reglas gramaticales que especifican cómo se pueden construir enunciados del lenguaje. Una producción es de la forma $\alpha \rightarrow \beta$, donde α y β son cadenas de terminales y no terminales. Una producción especifica que la cadena a puede transformarse en la cadena β. En el ejemplo anterior, (2.1) es el conjunto de producciones.
4. El *símbolo inicial* es un no terminal especial que comienza la generación de cualquier enunciado del lenguaje. En el ejemplo anterior, **enunciado** es el símbolo inicial.

Una vez dada una gramática, podemos generar los enunciados del lenguaje como sigue:

1. Comenzar con el símbolo inicial como la cadena actual de terminales y no terminales.
2. Si cualquier porción de la cadena actual de terminales y no terminales coincide con el lado izquierdo de una producción, remplazar esa porción de la cadena por el lado derecho de la producción. Para ser específico denote con a la cadena actual de terminales y no terminales. Más aún, suponga que es posible dividir **a** en tres subcadenas, α_1, α_2 y α_3, esto es, $\alpha = \alpha_1 \alpha_2 \alpha_3$. Si existe una producción $\alpha_2 \rightarrow \beta$, entonces podemos remplazar la subcadena cc$_2$ en **a** por p para obtener la cadena $\alpha_1\beta\alpha_3$. Usamos la notación $\alpha \Rightarrow \alpha_1\beta\alpha_3$ para indicar que la cadena **a** se pueda transformar en la cadena $\alpha_1\beta\alpha_3$. rem lazando una *porción* de la cadena **a** de acuerdo con una de las producciones de la gramática.
3. Cualquier cadena de terminales que se obtenga repitiendo el paso 2 es un enunciado del lenguaje. Observe que en el paso 2 existe la posibilidad de aplicar más de una producción para transformar la cadena actual de terminales y no terminales; en ese caso se puede escoger cualquiera de las producciones. Por otro lado, si en el paso 2 llegamos a una cadena de terminales y no terminales para la cual no se puede aplicar una producción, entonces hemos llegado a un callejón sin salida y debemos comenzar de nuevo con el símbolo inicial para obtener un enunciado en el lenguaje.

El proceso de generar un enunciado como el descrito también se conoce como *derivación.* En el ejemplo anterior, el enunciado *"un perro corre lentamente"* se deriva como sigue:

enunciado ⇒ **frase nominal frase-de-verbo-intransitivo**

⇒ **frase nominal verbo-intransitivo adverbio**

⇒ **frase nominal verbo-intransitivo** *rápidamente*

⇒ **frase nominal** *corre rápidamente*

⇒ **artículo sustantivo** *corre rápidamente*

⇒ **artículo** perro *corre rápidamente*

⇒ *un perro corre rápidamente*

Examinemos más ejemplos:

Ejemplo 2.4

Queremos construir una gramática para el lenguaje

$$L = \{aaaa, aabb, bbaa, bbbb\}$$

Como L tiene un número finito de cadenas, podemos simplemente listar todas las cadenas del lenguaje. Así, sea $T= \{a, b\}$ el conjunto de terminales, $N= \{S\}$ el conjunto de no terminales y S el símbolo inicial. Como conjunto de producciones tenemos

$$S \rightarrow aaaa$$
$$S \rightarrow aabb$$
$$S \rightarrow bbaa$$
$$S \rightarrow bbbb$$

Podemos, sin embargo, tener una gramática un poco más sencilla. Sea $N = \{S, A\}$ el conjunto de no terminales con S como símbolo inicial. El siguiente conjunto de producciones también especificará al lenguaje L:

$$S \rightarrow AA$$
$$A \rightarrow aa$$
$$A \rightarrow bb$$

❐

Ejemplo 2.5

Queremos construir una gramática para el lenguaje

$$L = \{a^i b^{2i} \mid i \geqq 1\}$$

Sea $T= \{a, b\}$ y $N = \{S\}$, con S como símbolo inicial. Sea

$$S \rightarrow aSbb$$
$$S \rightarrow abb$$

el conjunto de producciones. Entonces obtenemos, por ejemplo, la cadena *aaabbbbbb* como sigue:

$$S \Rightarrow aSbb \Rightarrow aaSbbbb \Rightarrow aaabbbbbb$$

❐

Ejemplo 2.6

Queremos construir una gramática para el lenguaje

$$L = \{x \mid x \in \{a, b\}^\star, \text{el número de letras } a \text{ en } x \text{ es múltiplo de 3}\}$$

Sean $T= \{a, b)$ y $N = \{S, A, B\}$, con S como símbolo inicial. El conjunto de producciones es

$$S \rightarrow bS$$
$$S \rightarrow b$$
$$S \rightarrow aA$$
$$A \rightarrow bA$$

$$A \to aB$$
$$B \to bB$$
$$B \to aS$$
$$B \to a$$

Ejemplo 2.7

Por ejemplo, tenemos

$$S \Rightarrow bS \Rightarrow bbS \Rightarrow bbaA \Rightarrow bbabA \to bbabaB \Rightarrow bbababB$$
$$\Rightarrow bbababbB \Rightarrow bbababbaS \Rightarrow bbababbab$$

Observamos que la no terminal *A* representa el conjunto de todas las cadenas en las que el número de letras *a* es $3k + 2$, y la no terminal *B* representa el conjunto de todas las cadenas en las que el número de letras *a* es $3k + 1$ para $k > 0$.

❐

Supongamos una determinada gramática en la cual $T= \{a, b\}$ y $N= \{S, A, B\}$, donde *S* es el símbolo inicial. Sea

$$S \to aB$$
$$S \to bA$$
$$A \to a$$
$$A \to aS$$
$$A \to bAA$$
$$B \to b$$
$$B \to bS$$
$$B \to aBB$$

el conjunto de producciones. Observamos que los enunciados del lenguaje son todas las cadenas de letras *a* y *b* en las que el número de letras *a* es igual al número de letras *b*. Dicha observación se aclara al distinguir que la no terminal *A* representa el conjunto de cadenas en las que el número de letras *a* es uno más que el número de letras *b, y* la no terminal *B* representa el conjunto de cadenas en las que el número de letras *b* es uno más que el número de letras *a*.

❐

Ejemplo 2.8

Queremos construir una gramática para el lenguaje

$$L = \{a^i b^j \mid i, j \geqq 1, i \neq j\}$$

Observamos que

$$L = L_1 \cup L_2$$

donde

$$L_1 = \{a^i b^j \mid i > j\}$$
$$L_2 = \{a^i b^j \mid i < j\}$$

Señalemos que

$$\begin{aligned} A &\to aA \\ A &\to aB \\ B &\to aBb \\ B &\to ab \end{aligned} \tag{2.2}$$

es un conjunto de producciones en una gramática para L_1 donde $\{a, b\}$ es el conjunto de terminales y $\{A, B\}$ es el conjunto de no terminales, con A como símbolo inicial. También

$$\begin{aligned} C &\to Cb \\ C &\to Db \\ D &\to aDb \\ D &\to ab \end{aligned} \tag{2.3}$$

es un conjunto de producciones en una gramática para L_2, donde $\{a, b\}$ es el conjunto de terminales y $\{C, D\}$ es el conjunto de no terminales, donde C es el símbolo inicial. Comprendemos de inmediato que al añadir las dos producciones

$$\begin{aligned} S &\to A \\ S &\to C \end{aligned}$$

a las producciones en (2.2) y (2.3) tenemos una gramática para L, con S como símbolo inicial. Sin embargo, podemos simplificar la gramática a

$$\begin{aligned} S &\to A \\ S &\to C \\ A &\to aA \\ A &\to aB \\ B &\to aBb \\ B &\to ab \\ C &\to Cb \\ C &\to Bb \end{aligned}$$

❐

Ejemplo 2.9 Consideremos la gramática siguiente, la cual especifica afirmaciones de asignación que incluyen identificadores, los operadores aritméticos + y * , el signo =, el paréntesis izquierdo, (, y el paréntesis derecho,). Sea *T = {A, B, C, D, +, ,, (,),=} y N =* **{af_de_asgn, exp, term, factor, id},** siendo **af_de_asgn** el símbolo inicial. Sea el siguiente el conjunto de producciones:

af_de_asgn → **id = exp**

exp → **exp + term**

exp → **term**

term → **term * factor**

term → **factor**

factor → **(exp)**

factor → **id**

id → ***A***

id → ***B***

id → ***C***

id → ***D***

Observemos que

af_de_asgn ⇒ **id = exp**

⇒ **id = exp + term**

⇒ **id = exp + term * factor**

⇒ **id = exp + term * (exp)**

⇒ **id = exp + term * (exp + term)**

⇒ **id = exp + term * (exp + factor)**

⇒ **id = exp + term * (exp + id)**

⇒ **id = exp + term * (exp + *B*)**

⇒ **id = exp + term * (term + *B*)**

⇒ **id = exp + term * (factor + *B*)**

⇒ **id = exp + term * (id + *B*)**

⇒ **id = exp + term * *(D + B)***

⇒ **id = exp + factor * *(D + B)***

⇒ **id = exp + id * (£> + *B*)**

⇒ **id = exp + *D* * *(D + B)***

$\Rightarrow$ **id** = **term** + $D * (D + B)$

$\Rightarrow$ **id** = **factor** + $D * (D + B)$

$\Rightarrow$ **id** = **id** + $D*(D + B)$

$\Rightarrow$ **id**$=A+D*(D + B)$

$\Rightarrow$ $C = A+D*(D + B)$

❐

2.6 TIPOS DE GRAMÁTICAS Y LENGUAJES

En la sección 2.5 vimos cómo se pueden usar gramáticas para especificar lenguajes. Pues resulta que usar gramáticas para especificar lenguajes nos conduce a una manera natural de clasificar lenguajes. Entrar en detalles al respecto rebasa el ámbito de este libro; de cualquier forma, sí queremos enunciar algunos de los resultados, sin demostración, de manera que el lector pueda comprender mejor el tema.

En adelante usaremos A y B para denotar no terminales arbitrarias, a *y* b para denotar terminales arbitrarias, y α y β para denotar cadenas arbitrarias de terminales y no terminales. Se dice que una gramática es una *gramática de tipo 3* si todas las producciones de la gramática son de las formas

$$\begin{aligned} A &\rightarrow a \\ A &\rightarrow aB \end{aligned} \tag{2.4}$$

o, de manera equivalente,† de las formas

$$\begin{aligned} A &\rightarrow a \\ A &\rightarrow Ba \end{aligned} \tag{2.5}$$

En otras palabras, en cualquier producción la cadena del lado izquierdo siempre es una sola no terminal y la cadena del lado derecho es una terminal o una terminal seguida de una no terminal. Así, la gramática del ejemplo 2.6 es una gramática de tipo 3.

---

† No es obvio que las formas en (2.4) sean equivalentes a las formas en (2.5) en el sentido de que los lenguajes que puedan especificarse usando producciones de las formas en (2.4) también se puedan especificar usando producciones de las formas en (2.5), y viceversa. Más aún, se puede demostrar que las producciones de las formas en (2.4) son equivalentes a producciones de las formas

$$\begin{aligned} A &\rightarrow \gamma \\ A &\rightarrow \gamma B \end{aligned}$$

y las producciones de las formas en (2.5) son equivalentes a las producciones de las formas

$$\begin{aligned} A &\rightarrow \gamma \\ A &\rightarrow B\gamma \end{aligned}$$

donde y es una cadena arbitraria de terminales. Vea los problemas 7.28 y 7.29.

En una *gramática de tipo 2,* toda la producción es de la forma

$$A \rightarrow \alpha$$

En otras palabras, en cualquier producción la cadena del lado izquierdo siempre es una sola no terminal. Es claro que una gramática de tipo 3 es trivialmente una gramática de tipo 2. Así, las gramáticas de los ejemplos 2.5 y 2.7 son gramáticas de tipo 2. En una *gramática de tipo 1,* para toda producción

$$\alpha \rightarrow \beta$$

la longitud de (3 es mayor o igual que la longitud de a. Por ejemplo, todas las producciones

$$A \rightarrow ab$$

$$A \rightarrow aA$$

$$aAb \rightarrow aBCb$$

satisfacen la condición, mientras que las producciones

$$aA \rightarrow a$$

$$ABc \rightarrow bc$$

no lo hacen. De nuevo, es claro que una gramática de tipo 3 o de tipo 2 también es, trivialmente, una gramática de tipo 1.

Una gramática de estructuras de frases como se definió antes, sin ninguna restricción, se conoce como *gramática de tipo 0.*

Hay diferentes tipos de lenguajes, correspondientes a diferentes tipos de gramáticas. Así, se dice que un lenguaje es de *tipo i (i* = 0, 1, 2, 3) si puede especificarse mediante una gramática de tipo / pero no puede especificarse por una gramática de tipo *(i* + 1). Por ejemplo, el lenguaje

$$L = \{a^k b^k \mid k \geqq 1\}$$

es un lenguaje de tipo 2, porque puede especificarse mediante la gramática de tipo 2

$$A \rightarrow aAb$$

$$A \rightarrow ab$$

Pero, por otro lado, *L* no se puede especificar mediante una gramática de tipo 3. (¿Cómo se puede demostrar que no existe una gramática de tipo 3 que especifique el lenguaje *Ll* Responderemos en el capítulo 7.) Así, correspondiendo a los cuatro tipos de gramáticas tenemos también cuatro tipo de lenguajes. De manera natural surgen algunas preguntas:

¿Hay lenguajes que no sean de tipo 0? La respuesta es afirmativa. En otras palabras, hay lenguajes que no se pueden especificar mediante gramáticas de estructuras de frases.

¿Qué sucede con todos los lenguajes de programación? Se pueden especificar mediante gramáticas de estructuras de frases. De hecho, todos son lenguajes (casi) de tipo 2.†

¿Toda clase de lenguajes de tipo i es no vacía? La respuesta es afirmativa pero no es tan obvia como lo parece. (Como en apariencia escogimos las definiciones de los varios tipos de gramáticas de manera arbitraria, no puede descartarse la posibilidad de que cualquier lenguaje que pueda especificarse mediante una gramática de tipo i también pueda especificarse mediante una gramática de tipo $(i + 1)$.)

Desde un punto de vista más práctico, lo importante es determinar cuándo una cadena dada pertenece, en efecto, a un lenguaje especificado mediante una gramática, y si así es, cómo se deriva esa cadena. En la sección 2.5 mostramos cómo derivar los enunciados de un lenguaje. Pero en la práctica, por ejemplo, cuando queremos construir un compilador para un lenguaje de programación, necesitamos determinar si una cadena dada es en efecto un enunciado válido del lenguaje. Después necesitamos descubrir cómo se derivó el enunciado de modo que podamos traducirlo (en código de instrucciones de máquina) de la misma manera. Todo esto, conceptualmente, se puede hacer mediante una búsqueda exhaustiva, pero se han desarrollado algoritmos eficientes para efectuar las tareas.

Muchos de estos temas se investigan en detalle en el estudio de la teoría de lenguajes formales y el diseño y construcción de compiladores.

2.7 NOTAS Y REFERENCIAS

Véase Stoll [10], en especial la sección 2.11 sobre paradojas de la teoría (intuitiva) de conjuntos. Véase también Smullyan [9]. El concepto de computabilidad es muy importante desde el punto de vista tanto de las ciencias de la computación como de las matemáticas. Para una introducción muy ilustrativa véase Minsky [8]. Véase Hennie [4] y Yasuhara [11] para un tratamiento más detallado. Como referencias generales sobre teoría de la computación y lenguajes formales, véase Harrison [3], Hopcroft y Ullman [5], Kfoury, Molí y Arbib [6], y Lewis y Papadimitriou [7]. El concepto de jerarquía de los lenguajes fue introducido por Chomsky [2]. Para una introducción al diseño y construcción de compiladores véase Ano y Ullman [1].

1. Aho, A. V. y J. D. Ullman: *Principles of Compiler Design,* Addison-Wesley Publishing Company, Reading, Mass., 1977.
2. Chomsky, N.: *On Certain Formal Properties of Grammars,* "Information and Control", 2, 137-167(1959).
3. Harrison, M. A.: *Introduction to Formal Language Theory,* Addison-Wesley Publishing Com pany, Reading, Mass., 1978.

† La palabra "casi" entre paréntesis merece una explicación. La mayoría de las características de los lenguajes de programación de alto nivel, como BASIC, FORTRAN y Pascal, se especifican mediante gramáticas de tipo 2. Sin embargo, algunas características no se pueden describir de esta manera. Como existe un cuerpo de conocimientos bien desarrollado acerca de las gramáticas y lenguajes de tipo 2, por lo común aplicamos muchas de las técnicas desarrolladas para los lenguajes de tipo 2 para manejar la mayor parte de las características y después se hacen algunos arreglos especiales para manejar las "peculiaridades".

4. Hennie, F. C: *Introduction to Computability,* Addison-Wesley Publishing Company, Reading, Mass., 1977.
5. Hopcroft, J. E. y J. D. Ullman: *Introduction to Automata Theory, Languages and Computation,* Addison-Wesley Publishing Company, Reading, Mass., 1979.
6. Kfoury, A. J., R. N. Moll y M. A. Arbib: *A Programming Approach to Computability,* Springer-Verlag, Nueva York, 1982.
7. Lewis, H. R. y C. H. Papadimitriou: *Elements of the Theory of Computation,* Prentice-Hall, Englewood Cliffs, N.J., 1981.
8. Minsky, M.: *Computation: Finite and Infinite Machines,* Prentice-Hall, Englewood Cliffs, N.J., 1967.
9. Smullyan, R.: *What Is the Name of This Book -The Riddle of Dracula and Other Logical Puzzles,* Prentice-Hall, Englewood Cliffs, N.J., 1978.
10. Stoll, R. R.: *Set Theory and Logic,* W. H. Freeman and Company, San Francisco, 1963.
11. Yasuhara, A.: *Recursive Function Theory and Logic,* Academic Press, Nueva York, 1971.

PROBLEMAS

2.1 El profesor Lai acaba de regresar de una visita a una isla donde cada habitante siempre dice la verdad o siempre miente. Nos dijo que oyó las siguientes afirmaciones hechas por dos de los habitantes de la isla, *A* y *B: A: B* siempre miente. *B: A* siempre dice la verdad. ¿Qué puede decir acerca de las vacaciones del profesor Lai?

2.2 Sean $A = \{a, b, c\}$, $B = \{b, c, d\}$ y

$$L_1 = \{a^i b^j \mid i \geq 1, j \geq 1\}$$
$$L_2 = \{b^i c^j \mid i \geq j \geq 1\}$$
$$L_3 = \{a^i b^j c^i d^j \mid i \geq 1, j \geq 1\}$$
$$L_4 = \{(ad)^i a^j d^j \mid i \geq 2, j \geq 1\}$$

Determine si cada una de las siguientes afirmaciones es verdadera o falsa.

a) L_1 es un lenguaje sobre A.
b) L_1 es un lenguaje sobre B.
c) L_2 es un lenguaje sobre $A \cup B$.
d) L_2 es un lenguaje sobre $A \cap B$.
e) L_3 es un lenguaje sobre $A \cup B$.
f) L_3 es un lenguaje sobre $A \cap B$.
g) L_4 es un lenguaje sobre $A \oplus B$.
h) L_1 es un lenguaje sobre $A - B$.
i) L_1 es un lenguaje sobre $B - A$.
j) $L_1 \cup L_2$ es un lenguaje sobre A.
k) $L_1 \cup L_2$ es un lenguaje sobre $A \cup B$.
l) $L_1 \cup L_2$ es un lenguaje sobre $A \cap B$.
m) $L_1 \cap L_2$ es un lenguaje sobre B.
n) $L_1 \cap L_2$ es un lenguaje sobre $A \cup B$.
o) $L_1 \cap L_2$ es un lenguaje sobre $A \cap B$.

2.3 Sean $A = \{a, b, c\}$, y $B = \{c, d\}$. Determine los siguientes conjuntos:

a) $\{a^i b^j d \mid i \geq 1, j \geq 0\} \cap A^\star$
b) $\{a^i b^j d \mid i \geq 1, j \geq 0\} \cap B^\star$

c) $\{(cd)^i(bc)^j \mid i \geqq 0, j \geqq 0\} \cap A^\bigstar$
d) $\{(cd)^i(bc)^j \mid i \geqq 0, j \geqq 0\} \cap B^\bigstar$

2.4 Obtenga expresiones más simples para los lenguajes especificados a continuación:
a) $\{xab \mid x \in \{a, b\}^i, i \geqq 3\} \cup \{xbb \mid x \in \{a, b\}^i, i \geq 3\}$
b) $\{a^i b^j \mid i, j = 1\} \cap \{a, b\}^\bigstar$
c) $\{(ab)^i c^i \mid i \geqq 1\} \cap \{(a)^i (bc)^i \mid i \geqq 1\}$
d) $\{a\}^\bigstar - \{a^i b^j a^i \mid i = 1, j \geqq 0\}$
e) $\{(ab)^i a(ba)^j \mid i, j = 1\} - \{a(ba)^i \mid i \geqq 1\}$
f) $\{a(ba)^i \mid i \geqq 1\} \oplus \{(ab)^i a \mid i \geqq 1\}$

2.5 Determine cuáles de los siguientes enunciados están en el lenguaje generado por la gramática especificada mediante las producciones en (2.1). En su caso, dé la derivación paso a paso.
a) *gato persigue el perro*
b) *el perro halla rápidamente*
c) *el gato halla el gato rápidamente*
d) *el gato halla lentamente*
e) *un perro persigue rápidamente*
f) *gato corre rápidamente*
g) *un gato lentamente persigue el perro*
h) *perro corre el gato*
i) *perro lentamente halla el gato*
j) *gato corre*

2.6 Determine si cada uno de los siguientes enunciados está en el lenguaje generado por la gramática dada en el ejemplo 2.9. De ser así, dé la derivación paso a paso.
a) $B = C * D + A$
b) $B = C + D*A$
c) $C = D + \{A + B)$
d) $A = (C*D) + B$
e) $B = (A+B*(C + D))$

2.7 Considere el lenguaje *L* especificado por la gramática *(T, N, S, P),* donde
$T = \{a, b, c\}$ es el conjunto de terminales.
$N = \{S, A, B\}$ es el conjunto de no terminales.
S es el símbolo especial.
$P = \{S$ -* $AB, A \to ab, A \to aAb, B \to c, B \to Be\}$ es el conjunto de producciones.
a) Determine si cada una de las siguientes cadenas es un enunciado del lenguaje.

aabb
aaabbc
aaabbbccc
ababee

b) Describa el lenguaje *L* en notación de teoría de conjuntos.

2.8 Considere una gramática en la cual *N* = **{entero_con_signo, signo, entero, dígito}** y *T*= {+, -, 0, 1} donde **entero_con_signo** es el símbolo inicial. Sea el conjunto de producciones

entero_con_signo → signo entero
signo → +
signo → -
entero → dígito entero

entero → dígito
dígito → 0
dígito → 1

Muestre una derivación de la cadena -010 del lenguaje.

2.9 Diseñe una gramática que especifique un lenguaje, incluyendo enunciados como los siguientes:

¿Tú entiendes?
¿Me gusta Juan?
¿El viene?
¿A ella le gusta María?

(Dejamos los detalles del lenguaje a la imaginación del lector.)

2.10 En lo que sigue, sea *{A, B, C, S}* el conjunto de no terminales, con *S* como símbolo inicial. Sea *{a, b, c}* el conjunto de terminales. Describa el lenguaje especificado por cada conjunto de

a) $\{S \to aA, S \to aS, A \to ab\}$
b) $\{S \to abS, S \to aA, A \to a\}$
c) $\{S \to aAB, A \to aB, A \to a, B \to b, B \to c\}$
d) $\{S \to aSA, S \to aB, A \to b, B \to c\}$
e) $\{S \to Sa, S \to AB, A \to aA, A \to a, B \to b\}$
f) $\{S \to aS, S \to b\}$
g) $\{S \to AB, A \to aA, A \to a, B \to Bb, B \to b\}$
h) $\{S \to AB, A \to ab, A \to aAb, B \to c, B \to Bc\}$
i) $\{S \to aS, S \to bA, A \to aA, A \to a\}$
j) $\{S \to aA, A \to bA, A \to bC, C \to cC, C \to c\}$
k) $\{S \to aAb, S \to bBa, A \to aAb, A \to c, B \to bBa, B \to c\}$
l) $\{S \to BA, A \to Aa, A \to a, B \to Bb, B \to c\}$
m) $\{S \to AB, S \to c, A \to aC, C \to bS, B \to aD, D \to b\}$
n) $\{S \to AB, A \to aA, A \to b, B \to bB, B \to a\}$
o) $\{S \to aaA, A \to aa, A \to aaA, A \to B, B \to b, B \to bB\}$

2.11 Dé una gramática que especifique a cada uno de los siguientes lenguajes:

a) $L = \{a^{2i}b^{2j} \mid i \geqq 1, j \geqq 1\}$
b) $L = \{(ab)^i c^{2j} \mid i \geqq 1, j \geqq 1\}$
c) $L = \{a^i b^j \mid i < j,\ j \geqq 1, j \geqq 1\}$
d) $L = \{a^i b^j \mid i \leqq j \leqq 2i, i \geqq 1\}$
e) $L = \{a^i b^i c^j \mid i \geqq 1, j \geqq 1\}$
f) $L = \{a^i b^j c^q \mid i + j = q, i \geqq 1, j \geqq 1\}$
g) $L = \{a^{2i} c b^{2j+1} \mid i \geqq 1, j \geqq 1\}$
h) $L = \{a^i b^i c^j d^j e^k \mid i > 0, j > 0, k > 0\}$

2.12 Dé una gramática que especifique cada uno de los lenguajes siguientes:

a) Todo enunciado en el lenguaje es una cadena de números iguales de letras *a* y letras *b*.
b) Todo enunciado en el lenguaje es una cadena de letras *a* y letras *b*, con el número de letras *a* siendo un múltiplo de 3.

2.13 Dé una gramática de tipo 2 que genere el lenguaje *L* que consiste de las cadenas de ceros y unos, con más ceros que unos.

2.14 Suponga que *L*, y L_2 son lenguajes de tipo 2.

a) Demuestre que $L_1 \cup L_2$ también es un lenguaje de tipo 2.
b) Demuestre que L_1L_2 también es un lenguaje de tipo 2 donde

$$L_1L_2 = \{\alpha\beta \mid \alpha \in L_1, \beta \in L_2\}$$

producciones, ya sea verbalmente o en notaciones de teoría de conjuntos.

2.15 Sea $\{A, B, C, S\}$ el conjunto de no terminales, con S como símbolo inicial. Sea $\{a, b)$ el conjunto de terminales. Para cada conjunto de producciones, determine el tipo de la gramática y el tipo del lenguaje correspondientes.

a) $\{S \to ABC, A \to a, A \to b, aB \to b, bB \to a, bC \to a, aC \to b\}$
b) $\{S \to AB, AB \to BA, A \to a, B \to b\}$
c) $\{S \to AB, S \to BA, A \to a, B \to b\}$
d) $\{S \to aB, S \to bA, A \to a, B \to b\}$

2.16 Para cada conjunto de producciones describa, ya sea verbalmente o en notaciones de teoría de conjuntos, el lenguaje especificado.

a) $\{S \to aSBC, S \to aBC, CB \to BC, aB \to ab, bB \to bb, bC \to bc, cC \to cc\}$
b) $\{S \to ABC, AB \to aAD, AB \to bAE, DC \to BaC, EC \to BbC, Da \to aD, Db \to bD, Ea \to aE, Eb \to bE, aB \to Ba, bB \to Bb, AB \to aF, AB \to bG, F0 \to 0F, F1 \to 1F, G0 \to 0G, G1 \to 1G, FC \to a, GC \to b\}$
c) $\{S \to aAB, Aa \to SBa, Ab \to SbB, B \to SA, B \to ba, aB \to b\}$

2.17 *a)* Dé una gramática de tipo 3 que genere el lenguaje

$$L = \{x \mid x \in \{a, b\}^{\star} \text{ y } x \text{ no contienen dos letras } a \text{ consecutivas}\}$$

b) Dé una gramática de tipo 2 que genere el lenguaje

$$L = \{x \mid x \in \{a, b\}^{\star} \text{ y } x \text{ contiene el doble de letras } a \text{ que de letras } b\}$$

CAPÍTULO

TRES

Permutaciones, combinaciones y probabilidad discreta

3.1 INTRODUCCIÓN

En la sección 1.6 discutimos algunos resultados sobre el tamaño de los conjuntos finitos. En este capítulo presentaremos resultados más amplios en esta dirección. Por ejemplo, si A es un conjunto finito de tamaño n, podemos desear conocer el número de los diferentes subconjuntos del conjunto A; esto es, el tamaño del conjunto potencia de A, $\mathscr{P}(A)$. Más aún, entre todos los subconjuntos de A, es probable que deseemos identificar el número de subconjuntos de tamaño k. También podríamos querer conocer el número de conjuntos ordenados, siendo los componentes de los conjuntos ordenados los elementos de A. Por ejemplo, si A es un conjunto de 10 senadores, el número de subconjuntos en $\mathscr{P}(A)$, que es igual a 2^{10}, es el número de comités diferentes que los senadores pueden formar [incluyendo el comité sin miembros, correspondiente al conjunto vacío en $\mathscr{P}(A)$]. Más aún, el número de subconjuntos de tamaño 6 en $\mathscr{P}(A)$, igual a 210, es el número de comités diferentes de 6 miembros que pueden formarse. Un conjunto ordenado de tamaño 3 con distintos componentes de A puede representar a los 3 más altos receptores de votos de entre los 10 senadores en la elección. Hay 720 de dichos conjuntos ordenados, que corresponden a los 720 resultados posibles diferentes. Un conjunto ordenado de 3 elementos con componentes de A no necesariamente distintos puede equivaler a los 3 representantes de 3 diferentes comités senatoriales consistentes en algunos de los 10 senadores. En este capítulo, discutiremos éstos y otros problemas relacionados en el contexto de permutaciones y combinaciones de objetos.

3.2 LAS REGIAS DE LA SUMA Y EL PRODUCTO

Por *experimento* entendemos el proceso físico que posee un número de resultados observables. Por ejemplo, colocar una pelota en una caja; colocar un número determinado de pelotas en un cierto número de cajas; seleccionar un estudiante representativo de entre un grupo de estudiantes; asignar oficinas a los profesores; hacer apuestas en una carrera de caballos; lanzar una moneda; tirar un par de dados, y jugar una mano de poker, son todos experimentos. En el experimento de colocar una pelota en una caja, sólo existe un resultado posible (sólo hay una manera de colocar una pelota en una caja); los dos resultados posibles de lanzar una moneda son *cara* y *cruz;* los seis resultados posibles de tirar un dado son 1, 2, 3, 4, 5 y 6, y los experimentos de jugar una mano de poker y seleccionar 5 estudiantes representativos entre 35 000 tienen muchos resultados posibles. Cuando consideremos los resultados posibles de varios experimentos, seguiremos las reglas establecidas a continuación:

Regla del producto. Si un experimento tiene m resultados posibles y otro experimento tiene n resultados posibles, entonces existen $m \times n$ resultados posibles cuando ambos experimentos tienen lugar.

Regla de la suma. Si un experimento tiene m resultados posibles y otro experimento tiene n resultados posibles, entonces existen $m + n$ resultados posibles cuando exactamente uno de estos experimentos tiene lugar.

Por ejemplo, si hay 52 maneras de seleccionar un representante para la clase de primer año y 49 maneras de seleccionar un representante para la clase de segundo año, entonces, de acuerdo con la regla del producto habrá 52 x 49 maneras de seleccionar a los representantes de ambas clases. Pero, de acuerdo con la regla de la suma, habrá 52 + 49 maneras de seleccionar a uno de los representantes. Como otro ejemplo, supongamos que se ofrecen siete cursos diferentes por la mañana y cinco por la tarde. Habrá 7 x 5 opciones para los estudiantes que deseen inscribirse en un curso por la mañana y otro por la tarde. Y habrá 7 + 5 opciones, si desean inscribirse en un solo curso.

3.3 PERMUTACIONES

Consideremos el sencillo problema de colocar tres pelotas de color rojo, azul y blanco en 10 cajas numeradas con 1,2, 3,..., 10. Deseamos conocer el número de maneras distintas en que las pelotas pueden ser colocadas en las cajas, si cada caja es capaz de contener sólo una pelota. Coloquemos las pelotas una a la vez, iniciando con la pelota roja, luego la pelota azul y después la blanca. Puesto que la pelota roja puede colocarse en cualquiera de las 10 cajas, la pelota azul en cualquiera de las nueve cajas restantes y la pelota blanca en cualquiera de las ocho restantes, el número total de maneras distintas de colocar estas pelotas es $10 \times 9 \times 8 = 720$.

El resultado de este ejemplo numérico puede ser generalizado de inmediato: suponga que se colocan r pelotas de distinto color en n cajas numeradas y diferentes, con la condición de que cada caja puede contener sólo una pelota. Debido a que la primera pelota puede ser colocada en cualquiera de las n cajas, la segunda pelota puede colocarse en cualquiera de las $(n - 1)$ cajas restantes,..., y la r-ésima pelota puede ser colocada en cualquiera de las $(n - r + 1)$ cajas restantes, el número total de maneras distintas de colocar las pelotas es

$$n(n-1)(n-2)\cdots(n-r+1)$$

lo cual también puede escribirse como†

$$\frac{n!}{(n-r)!}$$

Usamos la notación $P\{n, r)$ para la cantidad $n(n - 1)(n - 2) \ldots ($« - $r + 1)$.

Los ejemplos siguientes muestran que el problema de colocar pelotas en cajas no carece de interés como podría parecer.

Ejemplo 3.1

¿De cuántas maneras pueden ser programados tres exámenes dentro de un periodo de cinco días, de modo que el mismo día no sean programados dos exámenes? Si consideramos los tres exámenes como pelotas de distintos colores y los cinco días como cajas numeradas de manera diferente, obtenemos el resultado 5 x 4 x 3 = 60.

❒

Ejemplo 3-2

Suponga que tenemos siete habitaciones y queremos asignar cuatro de ellas, como oficinas, a cuatro programadores y usar las tres restantes para terminales de computadora. La asignación puede realizarse en 7 x 6 x 5 x 4 = 840 maneras diferentes, porque podemos ver el problema como el de colocar cuatro pelotas distintas (los programadores) dentro de siete cajas distintas (las habitaciones), siendo las tres cajas vacías restantes las habitaciones para las terminales de computadora (suponemos que los programadores son distintos, pero que todas las terminales de computadora son idénticas).

❒

Un experimento equivalente al de colocar pelotas en cajas es el de acomodar o permutar objetos distintos. Por permutar r de n objetos distintos, entendemos acomodar r de estos n objetos en algún orden. Por ejemplo, hay seis maneras de permutar dos de los tres objetos *a, b, c.* Ellas son *ab, ba, ac, ca, be y cb.* Puesto que para acomodar r de n objetos realizamos el llenado de r posiciones con r de los n objetos, hay n opciones para el objeto de la primera posición, $n - 1$ opciones para un objeto (de los $n - 1$ objetos restantes) de la segunda posición, ..., y $n - r + 1$ opciones para el objeto (de los $n - r + 1$ objetos restantes) de la r-ésima posición. Entonces, hay $n(n - 1) \ldots (n - r + 1)$ maneras de acomodar r de n objetos en

† $n!$ se lee "n factorial" y está definido como $n(n$ — l)$(n - 2) \ldots 2 \times 1$. También se ha convenido que 0! es igual a l.

orden.† En la terminología de conjuntos ordenados existen $n(n-1)\ldots(n-r+1)$ r-adas ordenadas que tienen componentes *distintas,* las cuales son elementos de un conjunto de tamaño *n.*

Consideremos los siguientes ejemplos:

Ejemplo 3.3

Determinemos el número de cifras decimales de cuatro dígitos que no contienen dígitos repetidos. Éste es un problema que consiste en acomodar 4 de los 10 dígitos 0, 1, 2,..., 9, y la respuesta es $P(10,4) = 5040$. Entre estos 5040 números, $9 \times 8 \times 7 = 504$ de ellos comienzan con 0. Por tanto, $5040 - 504 = 4536$ de ellos no comienzan con 0. El resultado también puede calcularse como

$$9 \times 9 \times 8 \times 7 = 4536$$

usando el argumento de que el primer dígito puede ser cualquiera de los nueve dígitos 1,2,..., 9, el segundo dígito puede ser cualquiera de los restantes *nueve* dígitos, y así sucesivamente.

❒

Ejemplo 3.4

El número de maneras en que podemos formar cadenas de cuatro letras distintas (del alfabeto inglés) seguidas por tres dígitos distintos es

$$P(26, 4) \times P(10, 3) = 258\,336\,000$$

❒

Volvamos al problema de colocar tres pelotas de colores diferentes dentro de 10 cajas con numeración distinta. Supongamos que una caja puede contener tantas pelotas como queramos. La pelota roja puede colocarse en cualquiera de las 10 cajas, como puede hacerse con la azul y con la blanca, por lo que el número total de maneras de colocación es

$$10 \times 10 \times 10 = 1000$$

En general, hay n^r maneras de colocar *r* pelotas de color dentro de *n* cajas numeradas, si una caja puede contener tantas pelotas como queramos. Consideremos otros ejemplos:

Ejemplo 3.5

Si programamos tres exámenes dentro de un periodo de cinco días sin restricción alguna sobre el número de exámenes programados cada día, el número total de maneras es $5^3 = 125$.

❒

Ejemplo 3.6

De un conjunto *A* determinemos el número de subconjuntos de tamaño *r.* Consideremos el problema de colocar *r* elementos de *A* en dos cajas. Así, en cada colocación definimos al subconjunto de *A* por los elementos colocados en la caja 1 y los elementos descartados los colocamos en la caja 2. Dado que hay 2^r maneras de colocar los *r* elementos, hay 2^r subconjuntos en $\mathscr{P}(A)$.

❒

† Un punto de vista diferente, que puede causar alguna confusión inicial pero que demostrará ser útil en el futuro, es considerar el problema como si se tratara de colocar pelotas en cajas. Al considerar *n* cajas correspondientes a los *n* objetos, y *r* pelotas correspondientes a las *r* posiciones en el arreglo, el colocamiento de una pelota en una cierta caja es equivalente a poner el objeto correspondiente a la caja en una posición correspondiente a la pelota en un arreglo. En consecuencia, el número de maneras de permutar *r* de *n* objetos es $P(n, r)$.

De modo similar, en términos de permutación de objetos decimos que si hay n tipos distintos de objetos con una fuente infinita de cada tipo, entonces hay n^r maneras de acomodar r de estos n tipos de objetos, porque hay n opciones del objeto para la primera posición, n opciones del objeto por la segunda posición,..., y n opciones del objeto para la r-ésima posición. Si usamos la terminología de conjuntos ordenados, entonces hay n^r r-adas ordenadas con sus componentes siendo elementos de un conjunto de tamaño n. Por ejemplo, hay 10^4 sucesiones decimales de cuatro dígitos. En consecuencia, $10^4 - P(10,4) = 4960$ que contienen uno o más dígitos repetidos.

Ejemplo 3.7

Primero observemos que hay 2^r sucesiones binarias de r-dígitos. En seguida nos preguntamos, de entre las 2^r sucesiones binarias de r-dígitos, ¿cuántas de ellas tienen un número par de unos? Podemos considerar parejas de sucesiones binarias tales que las dos sucesiones en cada par difieren sólo en el r-ésimo dígito. Es claro, una de las dos sucesiones en un par tiene un número par de unos y la otra tiene un número impar de unos. De esto se sigue que hay $\frac{1}{2} \cdot 2^r$ sucesiones binarias de r-dígitos que contienen un número par de unos.

Hay una manera un poco diferente de derivar el mismo resultado. Hay 2^{r-1} sucesiones binarias de $(r - 1)$-dígitos. Para una sucesión de $(r - 1)$-dígitos que tiene un número par de unos podemos agregar un 0 para obtener una sucesión de r-dígitos que tiene un número par de unos. Para una sucesión de $(r - 1)$-dígitos que tiene un número impar de unos, podemos agregar un 1 para obtener una sucesión de r-dígitos que tiene un número par de unos. Más aún, de estas dos maneras obtendremos todas las sucesiones de r-dígitos que tienen un número par de unos. Por tanto, hay 2^{r-1} de ellas. Dicha idea puede ser empleada para mejorar el desempeño de las computadoras, porque dentro de éstas, los datos son representados por sucesión de dígitos binarios. En el transcurso de la manipulación y transmisión de estas sucesiones binarias, se dice que hay un error si un 0 se convierte en un 1, o un 1 se convierte en 0. Para que los errores sean detectados debemos usar sucesiones binarias de $(r - 1)$-dígitos, para representar los datos y agregar un r-ésimo dígito a cada sucesión de modo que la sucesión resultante de r-dígitos siempre tenga un número par de unos. La aparición de un error (y por tanto, la aparición de un número impar de errores) dará origen a una sucesión binaria con un número impar de unos. La detección de una sucesión binaria con un número impar de unos significará la presencia de una condición de error.

Ahora nos preguntamos por el número de sucesiones quíntuples (sucesiones formadas con los dígitos 0, 1, 2, 3,4) de r-dígitos que contienen un número par de unos. Observemos que entre las 5^r sucesiones quíntuples de r-dígitos, hay 3^r de ellas que sólo contienen los dígitos 2, 3 y 4. Desde luego, estas sucesiones son, contadas como sucesiones con un número par de unos. Las restantes $5^r - 3^r$ pueden ser divididas en grupos de acuerdo con los patrones de dígitos 2, 3 y 4 en las sucesiones (por ejemplo, todas las de la forma *23xx344xx2xxx* estarán en un grupo, donde cada x es 0 o 1). Ya que la mitad de las sucesiones en cada grupo tiene un número par de unos, el número total de sucesiones quíntuples de r-dígitos con un número par de unos es $3^r + \frac{1}{2}(5^r - 3^r)$. □

Ejemplo 3.8

Supongamos que imprimimos todos los números de cinco dígitos sobre tiras de papel con un número en cada tira. Sin embargo, puesto que los dígitos 0, 1, 6, 8 y 9 se convierten en 0,1,9, 8 y 6 cuando se leen puestos de cabeza, hay pares de números que pueden compartir la misma tira sí ésta se lee correctamente o de cabeza. Por ejemplo, podemos formar una tira para los números 89166 y 99168. La pregunta es, ¿cuántas tiras diferentes tendremos que hacer para todos los números de cinco dígitos? Primero debemos observar que hay 10^5 números distintos de cinco dígitos, entre los cuales 5^5 pueden leerse ya sea derechos o de cabeza (son aquellos formados con los dígitos 0, 1, 6,8 y 9). No obstante, hay números que significan lo mismo ya sea derechos o de cabeza, por ejemplo, 16091, y hay $3(5^2)$ de tales números. (El dígito central de éstos debe ser 1, 0, u 8; además, el quinto dígito debe ser el primer dígito de cabeza, y el cuarto dígito debe ser el segundo dígito de cabeza.) En consecuencia, hay $5^5 - 3(5^2)$ números que pueden leerse ya sea derechos o de cabeza pero que se leerán de manera distinta. Estos números pueden ser divididos en parejas de modo que cada pareja de números comparta una tira. De esto se sigue que el número total de tiras diferentes que necesitamos es $10^5\text{-}[5^5\text{-}3(5^2)]/2$.

□

Ejemplo 3.9

Ahora demos un ejemplo sobre la aplicación del principio de inclusión y exclusión. Suponga que una estudiante desea elaborar una programación para un periodo de siete días, tiempo durante el cual ella estudiará una materia cada día. Ella cursa cuatro materias: matemáticas, física, química y economía. Es obvio, hay 4^7 programaciones diferentes. Pero deseamos conocer el número de programaciones que dedican al menos un día a cada materia.†

Si A_1 es el conjunto de programaciones en las cuales matemáticas nunca se incluye; A_2 denota el conjunto de programaciones en las cuales física nunca se incluye; A_3 incluye el conjunto de programaciones en las cuales química nunca se incluye, y A_4 denota el conjunto de programaciones en las cuales economía nunca se incluye, entonces

$$A_1 \cup A_2 \cup A_3 \cup A_4$$

es el conjunto de programaciones en las cuales una o más de las materias no se incluyen. Dado

$$|A_1| = |A_2| = |A_3| = |A_4| = 3^7$$

$$\begin{aligned} |A_1 \cap A_2| = |A_1 \cap A_3| &= |A_1 \cap A_4| = |A_2 \cap A_3| \\ &= |A_2 \cap A_4| = |A_3 \cap A_4| = 2^7 \end{aligned}$$

$$\begin{aligned} |A_1 \cap A_2 \cap A_3| = |A_1 \cap A_2 \cap A_4| &= |A_1 \cap A_3 \cap A_4| \\ &= |A_2 \cap A_3 \cap A_4| = 1^7 \end{aligned}$$

$$|A_1 \cap A_2 \cap A_3 \cap A_4| = 0$$

† Invitamos al lector a convencerse de que $P(7,4) \times 4^3$ no es la respuesta correcta. Hay un error en el argumento de que hay P(7,4) maneras de programar cuatro materias para cuatro de siete días, y 4^3 maneras de programar tres materias para los tres días restantes.

obtenemos

$$|A_1 \cup A_2 \cup A_3 \cup A_4| = 4(3^7) - 6(2^7) + 4$$

Por tanto, el número de programaciones en las cuales todas las materias estarán incluidas es

$$4^7 - 4(3^7) + 6(2^7) - 4$$

❒

Ahora consideremos el problema de colocar cuatro pelotas -dos rojas, una azul y una blanca-en 10 cajas numeradas. Repintamos las dos pelotas rojas con dos tonos de rojo, claro y oscuro, de manera que se diferencien. El número de maneras para colocar estas pelotas en las 10 cajas es $P(10,4) = 5040$. De entre estos 5040 arreglos, consideremos el arreglo en el cual la pelota en rojo claro está en la primera caja, la pelota en rojo oscuro está en la segunda caja, la pelota azul en la tercera caja y la pelota blanca en la cuarta caja. Si no distinguimos entre los dos tonos de rojo, es decir, si las dos pelotas rojas son indistinguibles, en realidad estas dos maneras de acomodarlas son una sola. Es cierto, los 5040 acomodamientos pueden ser acoplados en parejas en una forma similar, de manera que cada pareja de acomodamientos corresponde en realidad a un solo acomodamiento cuando los dos tonos de rojo no son distinguibles. En consecuencia, hay $\frac{5040}{2} = 2520$ maneras de acomodar dos pelotas rojas, una azul y una blanca en 10 cajas numeradas. Siguiendo el mismo argumento, vemos que el número de maneras de colocar tres pelotas rojas, una pelota azul y una blanca en 10 cajas numeradas es

$$\frac{P(10, 5)}{3!} = 5040$$

debido a que cada manera de colocar las tres pelotas rojas *indistinguibles*, una pelota azul y una pelota blanca corresponde a 3! maneras de colocar tres pelotas rojas *distinguibles*, una pelota azul y una pelota blanca.

En seguida deduciremos una fórmula general para el número de maneras de colocar r pelotas de color en n cajas numeradas, donde q_x de estas pelotas son de un color, q_2 de ellas son de un segundo color, . . . , y q_t de ellas son de un t-ésimo color. Observemos que un acomodamiento de las r pelotas no cambia al rearreglar las q_1 pelotas del mismo color entre las cajas en las cuales han sido colocadas, o al rearreglar las q_2 pelotas del mismo color entre las cajas en las cuales han sido colocadas,..., o al rearreglar las q_1 pelotas del mismo color entre las cajas en las cuales han sido colocadas. Por otro lado, si las r pelotas tuvieran distinto color, cualquier rearreglo originaría un acomodamiento diferente. De esto se sigue que cada forma de acomodar r pelotas, no todas de distinto color, corresponde a $q_1!\, q_2! \ldots q_t!$ maneras de acomodar r pelotas de distinto color. Debido a que hay *Pin, r)* maneras de acomodar r pelotas de distinto color en n cajas numeradas, el número total de maneras de acomodar r pelotas de color en n cajas numeradas, donde q_x son de un color, q_2 son de un segundo color, ..., y q_1 son de un t-ésimo color, es

$$\frac{P(n, r)}{q_1!\, q_2! \cdots q_r!} \tag{3.1}$$

Ejemplo 3.10 El número de maneras de pintar 12 oficinas de forma tal que 3 de ellas sean verdes, 2 rosas, 2 amarillas y las restantes blancas es

$$\frac{12!}{3!\,2!\,2!\,5!} = 166\,320$$

❒

En términos de arreglo de objetos, decimos que hay

$$\frac{n!}{q_1!\,q_2!\cdots q_t!} \tag{3.2}$$

maneras de acomodar n objetos, donde q_1 son de un tipo, q_2 son de un segundo tipo, ..., y q_1 son de un t-ésimo tipo. Ahora, si los n objetos fueran todos distintos, hay $n!$ maneras de acomodarlos. Por otro lado, en un arreglo de objetos que no son todos distintos la permutación de objetos del mismo tipo entre ellos mismos no cambiará el arreglo, obteniéndose así la fórmula (3.2).

Ejemplo 3.11 El número de mensajes diferentes que pueden representarse por una sucesión de tres líneas y dos puntos es

$$\frac{5!}{3!\,2!} = 10$$

❒

3.4 COMBINACIONES

Ahora consideremos el problema de colocar tres pelotas, todas ellas rojas, en 10 cajas que están numeradas 1,2, 3, ..., 10. Nuestro objetivo es conocer el número de maneras en que las pelotas pueden distribuirse, si cada caja puede contener sólo una pelota. De acuerdo con (3.1), la respuesta es

$$\frac{10 \times 9 \times 8}{3!}$$

En general, el número de maneras de distribuir r pelotas del mismo color en n cajas numeradas es

$$\frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

La cantidad $\frac{n!}{r!(n-r)!}$ se denota también por $C(n, r)$.

Examinemos los siguientes ejemplos:

Ejemplo 3.12 Un ama de casa desea programar cenas de espagueti tres veces por semana. Imaginemos las cenas de espagueti como tres pelotas y los siete días de la semana como siete cajas; entonces el número de maneras de programación es

$$\frac{7!}{3!\,4!} = 35$$

□

Ejemplo 3.13

Hay $C(32, 7)$ sucesiones binarias de longitud 32 y en cada una hay exactamente siete unos, podemos considerar el problema como la distribución de siete unos en 32 cajas numeradas (y entonces llenar las cajas vacías con ceros).

□

La selección de r objetos a partir de n objetos distintos es un problema equivalente a distribuir r pelotas indistinguibles en n cajas numeradas. Si vamos a seleccionar r de n objetos distintos, podemos imaginar los n objetos como cajas y marcar los objetos seleccionados con r marcadores idénticos, las pelotas. En consecuencia, el número de maneras de seleccionar r objetos de n objetos distintos es también $C(n, r)$. En otras palabras, para un conjunto de tamaño n se tienen $C(n, r)$ subconjuntos de tamaño r.

De acuerdo con la definición de $C(n, r)$, es claro que $C(n, r) = C(n, n - r)$. Existe un argumento combinatorio simple que confirma este resultado: seleccionar r objetos de n objetos es lo mismo que descartar los $n - r$ objetos que no serán seleccionados; entonces, tenemos que $C(n, r) = C(n, n - r)$.

Ejemplo 3.14

De entre 11 senadores hay $C(11, 5) = 462$ maneras de seleccionar un comité de 5 miembros. Más aún, hay $C(10, 4) = 210$ maneras de seleccionar un comité de cinco miembros de forma tal que un senador en particular, el senador A, siempre será incluido, y hay $C(10, 5) = 252$ maneras de seleccionar un comité de cinco miembros en forma tal que el senador A siempre será excluido. La pregunta es ¿de cuántas maneras podemos seleccionar un comité de cinco miembros para que al menos uno de los senadores A y B sea incluido? El número de selecciones que incluyen a ambos senadores, A y B, es $C(9, 3) = 84$. El número de selecciones que incluyen al senador A pero excluyen al senador B es $C(9, 4) = 126$, que es también el número de selecciones que incluyen al senador B pero excluyen al senador A. En consecuencia, el número total de maneras de realizar la selección es

$$84 + 126 + 126 = 336$$

De modo similar, ya que el número total de comités excluyendo a ambos senadores A y B es $C(9, 5)$, el número total de maneras de realizar la selección es

$$C(11, 5) - C(9, 5) = 462 - 126 = 336$$

El problema también tiene solución al aplicar el principio de inclusión y exclusión. De entre las 462 maneras de seleccionar 5 senadores, supongamos que A_1 y A_2 son los conjuntos de selecciones que incluyen al senador A y al senador B, respectivamente. Dado que

$$|A_1| = C(10, 4) = 210$$
$$|A_2| = C(10, 4) = 210$$
$$|A_1 \cap A_2| = C(9, 3) = 84$$

se sigue que

$$|A_1 \cup A_2| = 210 + 210 - 84 = 336$$

Ejemplo 3.15

A partir de un decágono convexo, consideremos todos los puntos de intersección, interiores al decágono, entre sólo dos diagonales, ¿en cuántos segmentos de línea serán divididas las diagonales por las intersecciones? Primero, tenemos que el número total de diagonales es igual a

$$C(10,2)-10 = 45-10 = 35$$

hay $C(10, 2)$ líneas rectas uniendo los $C(10, 2)$ pares de vértices, pero 10 de estas 45 líneas son los lados del decágono. Dado que cada cuatro vértices podemos contar una intersección entre diagonales, como se muestra en la figura 3.1 (el decágono es convexo), hay un total de $C(10,4) = 210$ intersecciones entre las diagonales (con sólo dos diagonales en el punto de intersección). Como una diagonal se divide en $k + 1$ segmentos de línea recta cuando hay ¿puntos de intersección sobre ella, y como cada punto de intersección se encuentra sobre dos diagonales, el número total de segmentos de línea recta en los cuales se dividen las diagonales es 35 + 2 x 210 = 455.

Suponga que vamos a distribuir r pelotas del mismo color en n cajas numeradas, y podemos colocar tantas pelotas como se quiera en una caja. El número de maneras de distribuir las pelotas es

$$\frac{(n+r-1)!}{r!(n-1)!} = C(n+r-1, r)$$

Una forma sencilla de verificar este resultado es considerar el problema de realizar un arreglo con $n + 1$ unos y r ceros con un 1 al inicio y un 1 al final de cada arreglo. Si consideramos los unos como las separaciones entre cajas y los ceros como pelotas, entonces cada arreglo corresponde a una manera de distribuir r pelotas del mismo color en n cajas numeradas. Por ejemplo, si $n = 5$ y $r = 4$, la sucesión

1011001101

puede verse como una distribución de cuatro pelotas en cinco cajas teniendo una pelota en la primera caja, ninguna pelota en la segunda caja, dos pelotas en la tercera caja, ninguna en la cuarta y una en la quinta. De acuerdo con (3.1), el número de maneras de arreglar r ceros y $n + 1$ unos con unos en ambos extremos de cada arreglo es

$$\frac{(n+r-1)!}{r!(n-1)!}$$

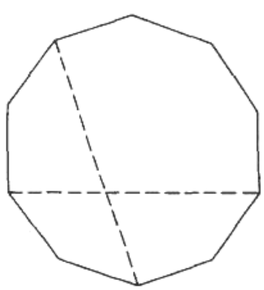


Figura 3.1

El problema de seleccionar r objetos a partir de n objetos distintos con selecciones repetidas, puede verse como aquel que utiliza r marcadores idénticos para marcar los n objetos distintos, donde cada objeto puede ser marcado con un número arbitrario de marcadores. Luego entonces, el número de maneras de seleccionar r objetos a partir de n objetos distintos, con selecciones repetidas es

$$\frac{(n+r-1)!}{r!(n-1)!} = C(n+r-1, r) \tag{3.3}$$

Ejemplo 3.17

Consideremos los siguientes ejemplos:

Ejemplo 3.16

El número de maneras de escoger tres de siete días (con repeticiones permitidas) es

$$C(7+3-1, 7) = C(9, 3) = 84$$

El número de maneras de escoger siete de tres días (con repetición necesariamente permitida) es

$$C(3+7-1, 7) = C(9, 7) = 36$$

❐

Una ficha de dominó está formada por dos cuadros, cada uno marcado con uno, dos, tres, cuatro, cinco o seis puntos, o se deja en blanco (sin puntos). Hay 28 fichas diferentes en un juego de dominó, debido a que el número de fichas distintas corresponde al número de maneras de seleccionar dos elementos a partir de los siete objetos "uno", "dos", "tres", "cuatro", "cinco", "seis" y "vacío" con repeticiones permitidas. Así, de acuerdo con (3.3), el número de fichas distintas es

$$C(7+2-1, 2) = C(8,2) = 28$$

❐

Ejemplo 3.18

Cuando se lanzan tres dados, el número de resultados diferentes es

$$C(6+3-1, 3) = C(8,3) = 56$$

debido a que lanzar tres dados equivale a seleccionar tres de los seis números 1,2,3,4, 5, 6, con repeticiones permitidas.

❐

Ejemplo 3.19

Nos preguntamos por el número de trayectorias diferentes para que una torre† se desplace, desde la esquina suroeste de un tablero de ajedrez hasta la esquina noroeste, sólo con movimientos en direcciones este y norte. Si permitimos que un 0 denote un paso en dirección este y que un 1 denote un paso en dirección norte, entonces el número de trayectorias es igual al número de maneras de formar arreglos con siete ceros y siete unos, el cual es

$$\frac{14!}{7!\,7!} = 3432$$

† Una torre es una pieza de ajedrez que puede moverse horizontal y verticalmente sobre el tablero de ajedrez.

Ahora nos preguntamos cuántas de estas trayectorias constan de cuatro movimientos hacia el este y tres movimientos hacia el norte (por un movimiento hacia el este entendemos un cierto número de pasos consecutivos en dirección este; un movimiento hacia el norte se define en forma análoga). El número de maneras de realizar una trayectoria con cuatro movimientos hacia el este es el mismo que el de distribuir siete pelotas indistinguibles en cuatro cajas distintas sin dejar cajas vacías. Colocamos una pelota en cada una de las cajas y luego distribuimos las tres restantes. Dado que el número de maneras para distribuir tres pelotas indistinguibles en cuatro cajas distintas con cada caja aceptando tantas pelotas como queramos es

$$C(4+3-1,3) = 20$$

el número de maneras para distribuir las siete pelotas en cuatro cajas, sin dejar alguna caja vacía, también es 20. De modo similar, el número de maneras de realizar una trayectoria con tres movimientos hacia el norte es

$$C(3+4-1,3) = 15$$

Luego entonces, la respuesta a nuestra pregunta es

$$20 \times 15 = 300$$

Ejemplo 3.20

Queremos determinar el número de maneras de sentar a cinco niños en una fila de 12 sillas. El problema es similar al de formar un arreglo con 12 objetos de seis tipos diferentes, siendo cada niño un objeto de tipo distinto y las siete sillas desocupadas objetos del mismo tipo. De acuerdo con (3.2), el número de arreglos es

$$\frac{12!}{7!}$$

Existe una forma alternativa de obtener el mismo resultado. Primero acomodamos a los cinco niños en una fila (hay 5! maneras de hacerlo), y después distribuimos, en forma arbitraria las siete sillas desocupadas, ya sea entre dos niños o entre los dos extremos. Ahora el problema de distribución consiste en acomodar siete pelotas del mismo color en seis cajas. Así, el número de maneras de hacer esto es

$$5! \times C(6+7-1,7) = 5! \times \frac{12!}{7!\,5!} = \frac{12!}{7!}$$

Suponga que deseamos sentar a los niños de tal forma que no queden dos niños sentados uno junto a otro. Invitamos al lector a verificar que el número de maneras de sentar a los niños es

$$5! \times C(6+3-1,3) = 5! \times \frac{8!}{3!\,5!} = \frac{8!}{3!}$$

Ejemplo 3-21

Queremos determinar el número de maneras de colocar $2t + 1$ pelotas indistinguibles en tres cajas distintas, de manera que dos cajas cualesquiera juntas contengan más pelotas que la tercera. El número total de maneras de colocar las pelotas sin considerar la restricción es

$$C(3 + 2t + 1 - 1, 2t + 1) = C(2t + 3, 2t + 1)$$

El número total de maneras de colocar las pelotas de modo que la primera caja contenga más pelotas que la segunda y terceras cajas combinadas es

$$C(3 + t - 1, t) = C(t + 2, t)$$

(Colocamos $t + 1$ pelotas en la primera caja y luego colocamos las t pelotas restantes en forma arbitraria en las tres cajas.) El mismo resultado se aplica al caso en que la segunda caja tiene más pelotas que la primera y tercera cajas combinadas, y al caso en que la tercera caja tiene más pelotas que la primera y segunda cajas combinadas. Así, la respuesta a nuestra pregunta es

$$\begin{aligned} C(2t + 3, 2t + 1) - 3C(t + 2, t) &= C(2t + 3, 2) - 3C(t + 2, 2) \\ &= \tfrac{1}{2}(2t + 3)(2t + 2) - \tfrac{3}{2}(t + 2)(t + 1) \\ &= \frac{t(t + 1)}{2} \end{aligned}$$

□

*3.5 GENERACIÓN DE PERMUTACIONES Y COMBINACIONES

Suponga que deseamos escribir las $n!$ permutaciones de n objetos distintos. Para $n = 3$, sólo hay seis permutaciones, y resulta trivial escribirlas. Para $n = 4$ hay 24 permutaciones y continúa siendo una tarea trivial mantener un registro y estar seguros de que escribimos todas las permutaciones sin omisiones ni repeticiones. Un problema interesante es encontrar procedimientos sistemáticos que generen exhaustivamente todas las permutaciones de n objetos. Para ilustrarlo presentaremos uno de estos procedimientos. Antes de describirlo, nos permitimos hacer una pregunta importante que puede surgir: si vamos a diseñar un procedimiento o a analizar un procedimiento que nos presentan, ¿cómo podemos estar seguros de que éste hará en realidad lo que se espera que haga, es decir, generar todas las permutaciones de n objetos? No existe manera alguna de probar en forma exhaustiva el procedimiento, ya que se supone que funcionará adecuadamente para todos los números enteros positivos n. Una manera de asegurarnos que el procedimiento generará en forma por demás exhaustiva todas las permutaciones sin repeticiones, es introduciendo un ordenamiento para las $n\backslash$ permutaciones. Si tenemos un procedimiento que genera una a una las permutaciones de acuerdo con este orden, entonces podemos estar seguros de que el procedimiento generará correctamente todas las permutaciones cuando inicie con la primera permutación en el orden y termine con la última permutación en el orden. Un orden que podemos usar es el *orden lexicográfico.* Sin pérdida de generalidad, permitamos que $\{1, 2, 3, \ldots, n\}$ sean los n objetos a permutar. Dadas dos permutaciones $a_1a_2 \ldots a_n$ y $b_1b_2 \ldots b_n$, diremos que $a_xa_2 \ldots a_n$ se encuentran antes que $b_1b_2 \ldots b_n$ en un orden lexicográfico si, para algún m con $1 \leq m < n$, $a_1 = b_1, a_2 = b_2, \ldots, a_{m-1} = b_{m-1}$ y $a_m < b_m$. Por ejemplo, la permutación

124635 se encuentra antes que la permutación 125643, y la permutación 125463 se encuentra después de la permutación 125346.

Supongamos que se nos da una permutación $a_1a_2 \ldots a_n$. Nuestra pregunta será ¿cuál es la siguiente permutación de acuerdo con el orden lexicográfico? No es difícil mostrar que la siguiente permutación $b_1b_2 \cdots b_n$ debe ser tal que:

1. $a_i = b_i$, $1 \leqq i \leqq m - 1$, y $a_m < b_m$, para el m más grande posible.
2. b_m es el elemento más pequeño de entre $a_{m+1}, a_{m+2}, \ldots, a_n$ tal que es mayor que a_m.
3. $b_{m+1} < b_{m+2} < \cdots < b_n$.

Por ejemplo, la permutación siguiente a 124653 en el orden lexicográfico es 125346. Para una permutación dada $a_1a_2 \ldots a_n$ notamos que el m más grande posible para el cual (1) se satisface es el m más grande posible para el cual a_m es menor que al menos una de entre $a_{m+1}, a_{m+2}, \ldots, a_n$. Un momento de reflexión muestra que éste también es el m más grande posible para el cual $a_m < a_{m+1}$.† Luego entonces, si examinamos la permutación $a_1a_2 \ldots a_m$ elemento por elemento y de *derecha a izquierda,* la primera vez que observamos una disminución, conocemos el valor de m y podemos determinar $b_mb_{m+1} \ldots b_m$ de acuerdo con (2) y (3). Por ejemplo, supongamos que nos han dado la permutación 124653. Cuando recorremos la permutación de derecha a izquierda y elemento por elemento, de acuerdo con (1), determinamos que la siguiente permutación es de la forma 12*xxxx.* En otras palabras, el subíndice m es igual a 3. Además, de acuerdo con (2), podemos determinar que la siguiente permutación es de la forma 125*xxx.* Por último, de acuerdo con (3), determinamos que la siguiente permutación es 125346 (en realidad, podemos usar el hecho de que $a_{m+1} > a_{m+2} > \ldots > a_n$ para llevar a cabo los pasos (2) y (3) de manera más sencilla. Al lector interesado se le sugiere ver el Prob. 3.55).

Nuestra observación nos conduce de inmediato a un procedimiento sistemático para la generación de las n! permutaciones de n objetos, comenzando con la permutación 1234... n y terminando con la permutación $n \ldots$ 4321. Más aún, sabemos que el procedimiento es efectivamente correcto. Invitamos al lector a convencerse por sí mismo de que nuestro procedimiento generará las permutaciones de los cuatro objetos 1,2,3,4 en el siguiente orden:

$$1234 \to 1243 \to 1324 \to 1342 \to 1423 \to 1432 \to 2134 \to 2143 \to$$
$$2314 \to 2341 \to 2413 \to 2431 \to 3124 \to 3142 \to 3214 \to 3241 \to$$
$$3412 \to 3421 \to 4123 \to 4132 \to 4213 \to 4231 \to 4312 \to 4321$$

Ahora, deseamos generar todos los k-subconjuntos‡ del conjunto {1,2, 3, ..., *n).* Para poder introducir un orden lexicográfico en los subconjuntos, primero convenimos que cada subconjunto será representado por una sucesión de sus elementos acomodados en un orden

† Suponga que m es el subíndice más grande posible para el cual a_m es menor que alguno de los $a_{m+1}, a_{m+2}, \ldots,$ A_n. No obstante, si $a_m > a_{m+1}$ y $a_m < a_{m+j}$ para algún $j > 1$, tenemos que $a_{m+1} < a_{m+j}$ contradiciendo la suposición de que m es el subíndice más grande posible para el cual a_m es menor que alguno de los $a_{m+1}, a_{m+2}, \ldots, a_n$. El argumento recíproco se deja al lector.

‡ Un subconjunto de tamaño k se abrevia como un k-subconjunto.

creciente. Podemos entonces formular secuencias de acuerdo con un orden lexicográfico.† Por ejemplo, los 4-subconjuntos de {1,2, 3,4, 5, 6} son representados y ordenados como

1234
1235
1236
1245
1246
1256
1345
1346
1356
1456
2345
2346
2356
2456
3456

Siguiendo con exactitud el mismo razonamiento que en el procedimiento para generar permutaciones, consideremos cómo podemos desarrollar un procedimiento para generar todos los &-subconjuntos del conjunto {1, 2, ..., n}. Sea $a_1a_2 \dots a_k$ un k-subconjunto. Podemos mostrar que el siguiente &-subconjunto $b_1b_2 \dots b_k$ de acuerdo con el orden lexicográfico debe ser tal que

1. $a_i = b_i$, $1 \leqq i \leqq m-1$, y $a_m < b_m$ para el m más grande posible.
2. $b_m = a_m + 1$.
3. $b_{j+1} = b_j + 1$ para $m \leqq j \leqq k-1$.

En la sucesión $a_1a_2 \dots a_k$, definimos el máximo valor posible de a_j como $n - k + j$. Así, el máximo valor posible de a_k es n; el máximo valor posible de a_{k-1}, es n - 1; el máximo valor posible de a_{k-2} es n - 2,..., y el máximo valor posible de a_1 es n — k + 1. Dado que en $a_1 a_2 \dots a_k$ el más grande m para el cual a_m no es igual a su máximo valor posible es el más grande m que satisface (1), podemos determinar m examinando $a_1 a_2 \dots a_k$ de *derecha a izquierda*, elemento por elemento. Una vez que el valor de m se determina es posible determinar $b_mb_{m+i}, \dots b_k$ de acuerdo con (2) y (3). Invitamos al lector a confirmar que los 4-subconjuntos de {1, 2, 3,4, 5, 6} mostrados con anterioridad fueron en efecto generados por nuestro procedimiento.

3.6 PROBABILIDAD DISCRETA

Como un ejemplo de la aplicación de algunos de los conceptos y herramientas estudiados en el capítulo 1 y en este capítulo, presentamos una breve introducción a la teoría de probabilidad discreta. Recordemos que un *experimento* es un proceso físico que tiene un

† Para dos k-subconjuntos $a_1a_2 \cdots a_k$ y $b_1b_2 \cdots b_k$, diremos que $a_1a_2 \cdots a_k$ se encuentra antes de $b_1b_2 \cdots b_k$ en un orden lexicográfico si para algún $1 \leqq m \leqq k$, $a_1 = b_1$, $a_2 = b_2$, ..., $a_{m-1} = b_{m-1}$ y $a_m < b_m$.

número de resultados observables. En secciones previas de este capítulo, estudiamos diferentes maneras de calcular el número de resultados de un experimento. En los ejemplos encontramos que jugar una mano de póker tiene C(52, 5) = 2 598 960 posibles resultados; y examinar la boleta de calificaciones de un estudiante tiene 5^4 posibles resultados (suponiendo que el estudiante toma cuatro cursos, y las cinco posibles calificaciones son *A, B, C, D, F).* En nuestro modelo de un proceso físico, estos resultados se consideran *mutuamente exclusivos* y *exhaustivos;* esto es, exactamente un resultado tendrá lugar en cualquier caso particular del experimento. Así, cuando lanzamos una moneda, se observa una cara o una cruz. No es posible que se observen ambos resultados, y tampoco es posible que no se observe ninguno de ellos (si en realidad creemos que una moneda podrá detenerse sobre su canto, deberemos incluir en nuestro modelo tres posibles resultados cuando se lanza una moneda, a saber, *cara, cruz* y *cantó).*

En un sentido formal, nos referimos al conjunto de todos los resultados posibles de un experimento como al *espacio muestral* del experimento. También nos referimos a los resultados en un espacio muestral como a las *muestras* o *puntos muéstrales.* Utilizaremos la notación $S = \{x_1, x_2, . . . x_i, ...\}$ para un espacio muestral *S*, constituido por los puntos muéstrales $x_1 x_2, \ldots, x_i$, y así sucesivamente. Un espacio muestral que tiene un número finito o infinito *numerable* de puntos muéstrales es llamado un *espacio muestral discreto.* Restringiremos nuestra discusión a espacios muéstrales discretos. (Una discusión más profunda de espacios muéstrales con un número infinito *no-numerable* de puntos muéstrales requeriría algunos conceptos y herramientas avanzadas de análisis matemático.) Por ejemplo, en el experimento de lanzar una moneda, el espacio muestral es un conjunto $S = \{c, z\}$, el cual consiste de los dos posibles resultados *c* (cara) y *z* (cruz). Para el experimento de lanzar dos monedas, el espacio muestral es un conjunto $S = \{cc, cz, zc, zz\}$, el cual consiste en los cuatro resultados posibles cara-cara, cara-cruz, cruz-cara y cruz-cruz. En el experimento de esperar la llegada de un autobús a la parada de autobuses, el espacio muestral es un conjunto $S = \{0, 1, 2, 3, ..., 30\}$ en el cual los resultados son los tiempos de espera en un intervalo de 0 a 30 minutos. Consideremos el experimento de disparar hacia un blanco hasta que haya un acierto. El espacio muestral es un conjunto infinito numerable $S = \{a, na, nna, nnna, ... \}$, donde *a* denota un acierto, *na* denota un no-acierto seguido de un acierto, *nna* denota dos no-aciertos seguidos por un acierto, etcétera.

Asociado con cada punto muestral en un espacio muestral hay un número real llamado la *probabilidad* de ese punto muestral. En el punto muestral x_i usaremos $p(x_i)$ para denotar la probabilidad asociada con x_i. Las probabilidades asociadas con los puntos muéstrales deben satisfacer:

1. La probabilidad de cada punto muestral es un número no negativo menor o igual a 1. Esto es, para cada x_i en *S*, $0 \leq p(x_i) \leq 1$.
2. La suma de las probabilidades de todos los puntos muéstrales en el espacio muestral es igual a 1. Esto es, $\Sigma_{x_i \in S}\, p(x_i) = 1$.

La probabilidad de un punto muestral es una medida de la posibilidad de ocurrencia de ese punto. Un punto muestral con mayor probabilidad es más factible que suceda, en tanto que un punto muestral con probabilidad más pequeña es menos factible que ocurra. En un sentido cuantitativo, si realizamos un experimento un gran número de veces, la probabilidad de un punto muestral es una medida de la fracción de veces en las cuales el resultado particular sucede. Por ejemplo, en el espacio muestral del experimento de lanzar una moneda no cargada, la probabilidad del resultado *cara* es $\frac{1}{2}$, y la probabilidad del resultado *cruz* también es $\frac{1}{2}$. Así, si lanzamos la moneda muchas veces, aproximadamente la mitad de los resultados serán *caras* y la mitad serán *cruces*. Por otro lado, en el espacio muestral del experimento de lanzar una moneda cargada, la probabilidad del resultado *cara* podría ser $\frac{2}{3}$, en tanto que la probabilidad del resultado *cruz* podría ser $\frac{1}{3}$. En tal caso, cuando lanzamos muchas veces la moneda, aproximadamente dos tercios de los resultados serán *cara*, y aproximadamente un tercio de los resultados serán *cruz*. Además, para el experimento de lanzar dos monedas el espacio muestral es $S = \{cc, cz, zc, zz\}$, y tenemos

$$p(cc) = \frac{1}{4}$$

$$p(cz) = \frac{1}{4}$$

$$p(zc) = \frac{1}{4}$$

$$p(zz) = \frac{1}{4}$$

si la moneda no está cargada. O podríamos tener

$$p(cc) = \frac{4}{9}$$

$$p(cz) = \frac{2}{9}$$

$$p(zc) = \frac{2}{9}$$

$$p(zz) = \frac{1}{9}$$

si la moneda está cargada. Para el experimento de disparar hacia un blanco hasta que haya un acierto, el espacio muestral es $S = \{a, na, nna, nnna, \ldots\}$, y tenemos que

$$p(a) = \frac{1}{2}$$

$$p(na) = \frac{1}{4}$$

$$p(nna) = \frac{1}{8}$$

$$\cdots\cdots\cdots$$

$$p(\underbrace{nn\ldots n}_{k}a) = 2^{-(k+1)}$$

Además, en el espacio muestral de un experimento cualquiera, un punto muestral con probabilidad 1 corresponde a un resultado que tendrá lugar con seguridad; un punto muestral con probabilidad 0 corresponde a un resultado que nunca sucederá.

Ahora podemos explicar la importancia física de las dos condiciones establecidas con anterioridad, acerca de las probabilidades asociadas con los puntos muéstrales en un espacio muestral. Desde luego, si la probabilidad asociada con un punto muestral es una medida de la frecuencia de ocurrencia del resultado de un experimento, carece de sentido suponer para ésta un valor negativo o suponer un valor mayor que 1. Además, dado que el espacio muestral contiene todos los resultados posibles de un experimento, la suma de probabilidades de los puntos muéstrales debe ser exactamente igual a 1.

Supongamos que nos proporcionan las probabilidades de los resultados de un experimento, bien sea basadas en datos estadísticos o sencillamente en la *estimación intuitiva* de alguien.

Un *evento* es un subconjunto de resultados de un experimento. Se dice que un evento ocurre si cualquiera de los puntos muéstrales en el evento ocurre. Así, cuando tiramos un dado, obtener un 1 es un evento, obtener un número impar (1, 3 o 5) es otro evento. Un evento que contiene un punto muestral se denomina evento *simple,* y un evento que contiene más de un punto muestral se denomina evento *compuesto.* La probabilidad de ocurrencia de un evento se define como la suma de las probabilidades de los puntos muéstrales en el subconjunto. Dado que los puntos muéstrales son resultados mutuamente exclusivos de un experimento, la probabilidad de un evento es una medida de la frecuencia de ocurrencia del evento. Así, en notación de teoría de conjuntos, un evento A es un subconjunto del espacio muestral S. La probabilidad del evento A, denotado $p(A)$, es igual a $\Sigma_{x_i \in A}\, p(x_i)$.

Examinemos algunos ejemplo ilustrativos:

Ejemplo 3.22

Para el experimento de tirar un dado, el espacio muestral consiste en seis puntos muéstrales. Si suponemos que la probabilidad de ocurrencia de cada uno es $\frac{1}{6}$, entonces la probabilidad de obtener un número impar es igual a

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$$

Suponga que tenemos un dado "cargado" y que la probabilidad de obtener un 1 es $\frac{1}{3}$ y la probabilidad de obtener cada uno de los restantes números es ^. Luego entonces, la probabilidad de obtener un número impar es

$$\frac{1}{3}+\frac{2}{15}+\frac{2}{15}=\frac{3}{5}$$

y la probabilidad de obtener un número par es

$$\frac{2}{15}+\frac{2}{15}+\frac{2}{15}=\frac{2}{5}$$

❒

Ejemplo 3.23

Consideremos el problema de tratar una mano de póker a partir de una baraja de 52 cartas. El espacio muestral consiste en $C(52, 5)$ puntos muéstrales correspondientes a las $C(52, 5)$ manos diferentes que se pueden obtener. Supongamos que estos resultados tienen probabilidades iguales; esto es, la probabilidad de que se obtenga una mano en particular es igual a 1/C(52, 5). Para determinar la probabilidad de obtener cuatro ases, observemos que 48 de los C(52, 5) posibles resultados contienen cuatro ases; así, la probabilidad es 48/C(52, 5) = 0.0000185.

❐

Ejemplo 3.24

Confirmaremos la aseveración de que en un grupo de 23 personas la posibilidad de que haya entre ellas dos con el mismo día de nacimiento, es menor que un 50-50. El espacio muestral consiste en 366^{23} puntos muéstrales que corresponden a todas las distribuciones posibles de los días de nacimiento de 23 personas. Supongamos que estas distribuciones son equiprobables. Dado que *P(366*, 23) de las 366^{23} muestras corresponden a distribuciones de días de nacimiento tales que no hay dos de las 23 personas con el mismo día de nacimiento, la probabilidad de que no haya dos personas con el mismo día de nacimiento es

$$\frac{p(366, 23)}{366^{23}} = 0.494$$

❐

Ejemplo 3.25

Ocho estudiantes esperan en fila para tener una entrevista. Queremos determinar la probabilidad de que en la fila haya dos de primer año, dos de segundo, dos de tercero y dos de cuarto. El espacio muestral consiste en 4^8 puntos muéstrales que corresponden a todas las posibilidades de grados de donde pueden provenir los estudiantes. Supongan que estos puntos muéstrales son equiprobables. Hay $8!/2!\,2!\,2!\,2!$ puntos muéstrales que corresponden al caso en que haya dos estudiantes de cada grado. Así, la probabilidad es

$$\frac{8!}{2!\,2!\,2!\,2!\,4^8} = 0.0385$$

❐

Ejemplo 3.26

Para el experimento de dispararle a un blanco hasta que haya un acierto, suponemos la probabilidad de ocurrencia de la muestra que tiene *k* no-aciertos antes de un acierto como 2-($^{k+1}$). Si *A* denota el evento de que haya un acierto antes de no más de 5 no-aciertos, entonces *A* = *{a, na, nna, nnna, nnnna, nnnnná)* y

$$p(A) = \sum_{k=0}^{5} 2^{-(k+1)} = 0.984$$

Si *B* denota el evento de que haya un acierto después de un número impar de no-aciertos, entonces

Además, si *C* denota el evento de que haya un acierto después de un número par de no-aciertos (incluyendo aciertos). Entonces

$$p(C) = \sum_{i=0}^{\infty} 2^{-2i+1} = \frac{2}{3}$$

❐

Una vez más observaremos cómo los conceptos elementales de teoría de conjuntos nos permiten introducir precisa y consistentemente nuevas definiciones. Dados dos eventos *A* y

B, el evento de que ambos *A y B* ocurran corresponde al conjunto de puntos muéstrales $A \cap B$. Usaremos $A \cap B$ para denotar tal evento. Más aún, la probabilidad de ocurrencia del evento, denotada por $p(A \cap B)$, es igual a $\Sigma_{x_i \in A \cap B}\, p(x_i)$. Asimismo, dados dos eventos *A* y *B*, el evento de que ocurra *A* o *B* o ambos corresponde al conjunto de puntos muéstrales $A \cup B$. Además, el evento de que ocurra *A* pero no *B*, corresponde al conjunto de puntos muéstrales *A* - *B;* el evento de que ocurra uno pero no ambos corresponde al conjunto de puntos muéstrales $A \oplus B$. Desde luego, estos eventos son denotados por $A \cup B$, $A - B$, y $A \oplus B$, y sus probabilidades correspondientes pueden calcularse como $\Sigma_{x_i \in A \cup B}\, p(x_i)$, $\Sigma_{x_i \in A - B}\, p(x_i)$, $\Sigma_{x_i \in A \oplus B}\, p(x_i)$, respectivamente.

Ahora veremos algunos ejemplos:

Ejemplo 3.27

Los datos digitales recibidos desde un sitio remoto podrían llenar de 0 a 32 *buffers.* El espacio muestral es *S*= {0, 1,2,..., 32}, donde el punto muestral *i* denota que *i* de los *buffers* están llenos. Se tiene entonces que

$$p(i) = \frac{1}{561}(33 - i)$$

Si *A* denota el evento de que a lo más 16 *buffers* están llenos, y *B* denota el evento de que un número impar de *buffers* están llenos, entonces

$$\begin{aligned} A &= \{0, 1, 2, \ldots, 16\} \\ B &= \{1, 3, 5, \ldots, 31\} \\ A \cap B &= \{1, 3, 5, \ldots, 15\} \end{aligned}$$

y

$$p(A) = \frac{1}{561}\sum_{i=0}^{16}(33 - i) = \frac{425}{561} = 0.758$$

$$p(B) = \frac{1}{561}\sum_{\substack{i=1 \\ i \text{ impar}}}^{31}(33 - i) = \frac{272}{561} = 0.485$$

$$p(A \cap B) = \frac{1}{561}\sum_{\substack{i=1 \\ i \text{ impar}}}^{15}(33 - i) = \frac{200}{561} = 0.357$$

❐

Ejemplo 3-28

De entre 100 000 personas, 51 500 son mujeres y 48 500 hombres. Entre las mujeres 9 000 son calvas, y entre los hombres 30 200 son calvos. Suponga que escogemos una persona al azar. Tendremos *S* = *{mc, mn, hc, hn}* como espacio muestral, con *mc* denotando una mujer calva, *mn* una mujer no calva, *he* un hombre calvo, y *hn* un hombre no calvo. Así, tenemos

$$\begin{aligned} p(mc) &= 0.090 \\ p(mn) &= 0.425 \\ p(hc) &= 0.302 \\ p(hn) &= 0.183 \end{aligned}$$

Si A denota el evento de que una persona calva fuese escogida, y B el evento de que una mujer fuese escogida. Entonces $A \cap B$ es el evento de que uña mujer calva fuese escogida, $A \cup B$ el evento de que una persona calva o una mujer fuese escogida, $A \oplus B$ el evento de que una mujer con cabello o un hombre calvo fuese escogido, y $B - A$ el evento de que una mujer con cabello fuese escogida. Así, tenemos

$$\begin{aligned}
p(A) &= 0.90 + 0.302 = 0.392\\
p(B) &= 0.090 + 0.425 = 0.515\\
p(A \cap B) &= 0.090\\
p(A \cup B) &= 0.090 + 0.425 + 0.302 = 0.817\\
p(A \oplus B) &= 0.425 + 0.302 = 0.727\\
p(B - A) &= 0.425
\end{aligned}$$

❒

Ejemplo 3.29

Diez hombres llegaron a una fiesta y depositaron sus sombreros en el guardarropa. Los sombreros les fueron devueltos en forma aleatoria al retirarse. Queremos saber la probabilidad de que ningún hombre haya recibido su propio sombrero de vuelta. Para el experimento de regresar los sombreros a los hombres, el espacio muestral consiste de 10! puntos muéstrales correspondientes a las 10! permutaciones posibles de los sombreros. Supongamos que cada permutación ocurre con igual probabilidad, esto es, 1/10! En consecuencia, la probabilidad de que ningún hombre reciba su propio sombrero es igual a 1/10! veces el número de permutaciones en las cuales ningún hombre recibe su propio sombrero. Si A denota el conjunto de puntos muéstrales en los cuales el i-ésimo hombre recibe su propio sombrero, el lector puede confirmar que, usando el principio de inclusión y exclusión, obtenemos

$$|A_1 \cup A_2 \cup \cdots \cup A_{10}| = \binom{10}{1}9! - \binom{10}{2}8! + \binom{10}{3}7! - \cdots + \binom{10}{9}1! - \binom{10}{10}0!$$

Por tanto, la probabilidad de que ningún hombre reciba su propio sombrero es:

$$\frac{1}{10}\left[\ 10! - \binom{10}{1}9! + \binom{10}{2}8! - \binom{10}{3}7! + \cdots - \binom{10}{9}1! + \binom{10}{10}0!\ \right]$$

$$= 1 - \frac{1}{1!} + \frac{2}{2!} - \frac{3}{3!} + \cdots - \frac{9}{9!} + \frac{10}{10!} = 0.36788$$

Es posible que en forma intuitiva no habríamos adivinado que la probabilidad podía resultar tan grande.

❒

*3.7 PROBABILIDAD CONDICIONAL

Suponga que se lanza un dado y que deseamos conocer la probabilidad de que el resultado sea 4. Suponemos que los seis resultados son equiprobables. Es obvio, la respuesta es un sexto. Ahora, suponga que se tira un dado, y nos informan que el número fue par. De nuevo, deseamos conocer la probabilidad de que el resultado sea 4. Nos damos cuenta de que sólo 2,4 o 6 son los resultados posibles, la probabilidad que tiene el 4 de aparecer es mayor que un sexto. En efecto, quizá el lector ha llegado a la respuesta: la probabilidad de aparecer que tiene el 4 es un tercio.

Consideremos nuevamente el problema de las personas calvas en el ejemplo 3.28. Suponga que escogemos una persona al azar. Como se mostró antes, la probabilidad de que la persona sea calva es 0.392. Supongamos que nos informaron que la persona fue una mujer. Entonces podríamos decir, al menos intuitivamente, que la probabilidad de que esta persona haya sido calva podría ser menor que 0.392. Por otro lado, si se nos informa que la persona fue hombre, entonces la probabilidad de que éste sea calvo podría ser mayor que 0.392.

Estos dos ejemplos remiten a la noción de probabilidad condicional de un evento. Si *S* es un espacio muestral y *A* y *B* son dos eventos en *S,* la probabilidad de que ocurra el evento *A* dado que el evento *B* ha ocurrido, se define como *Improbabilidad condicional* del evento *A* dada la ocurrencia de *B,* la cual se denota por *p(A \ B).* En el ejemplo de tirar un dado, si *A* denota el evento: "el resultado es 4", y *B* denota el evento: "el resultado es un número par", la probabilidad condicional *p(A | B)* es igual a j. En el ejemplo de las personas calvas, si *A* denota el evento de que se escogiera a una persona calva, *B* denota el evento de que se escogiera una mujer, y *C* denota el evento de que se escogiera un hombre, podríamos estar de acuerdo en que *p(A \ B)* es menor que *p{A),* como ya se dijo. En otras palabras, puesto que fue escogida una mujer, existe una menor oportunidad de que una persona calva sea escogida. No obstante *p(A* | C) es mayor que *p(A),* dado que un hombre fue escogido, las oportunidades de que una persona calva sea escogida son mayores. Además, aun cuando la probabilidad *p(B)* de que una mujer haya sido escogida es mayor que la probabilidad *p(C)* de que un hombre haya sido escogido, la probabilidad condicional *p(B \ A)* de que una mujer haya sido escogida ya que una persona calva fue escogida es menor que la probabilidad condicional *p(C \ A)* de que haya sido escogido un hombre dado que una persona calva fue escogida. Más adelante mostraremos cómo calcular estas probabilidades condicionales para confirmar todas estas nociones intuitivas.

La ocurrencia del evento *B* cambia efectivamente las probabilidades asociadas con los puntos muéstrales en el espacio muestral. Es obvio que, la probabilidad asociada con un punto muestral, no incluido en el evento *B,* se convierte en 0. Por otro lado, la probabilidad asociada con un punto muestral incluido en el evento *B* se incrementa. Examinemos de nuevo el ejemplo de tirar un dado. Si nos informan que ha aparecido un número par, las probabilidades de los puntos muéstrales 1,3 y 5 se convierten en 0, debido a que ninguno de ellos podría haber ocurrido. Por otro lado, las probabilidades de los puntos muéstrales 2,4 y 6 se convierten en un tercio (suponiendo que todos los resultados posibles son igualmente probables). Así, en efecto,

$$p(\text{aparezca 4} \mid \text{ha aparecido un número par}) = \frac{1}{3}$$

En general, $p_B(x_i)$ denota la probabilidad asociada con el punto muestral x_i, dado que el evento *B* ha ocurrido. Como señalamos, para $x_i \notin B$, $p_B(x_i) = 0$.Sin embargo, para los puntos muéstrales en el evento *B* sus frecuencias *relativas* de ocurrencia permanecen iguales en tanto que la suma de sus probabilidades debe ser igual a 1, esto es, $\Sigma_{x_i \in B}\, p_B(x_i) = 1$.

En consecuencia, necesitamos escalar la probabilidad de cada uno de estos puntos de $p(x_i)$ a $p(x_i)/p(B)$. Así, tenemos

$$p_B(x_i) = \begin{cases} 0 & x_i \notin B \\ \dfrac{p(x_i)}{p(B)} & x_i \in B \end{cases}$$

De esto se sigue que

$$\begin{aligned} p(A \mid B) &= \sum_{x_i \in A \cap B} p_B(x_i) \\ &= \sum_{x_i \in A \cap B} \frac{p_B(x_i)}{p(B)} \\ &= \frac{1}{p(B)} \sum_{x_i \in A \cap B} p_B(x_i) \\ &= \frac{p(A \cap B)}{p(B)} \end{aligned}$$

Ejemplo 3.30

Para el ejemplo de escoger una persona al azar, si *A* denota el evento de que una persona calva fue escogida, *B* el evento de que una mujer fue escogida, y *C* el evento de que un hombre fue escogido, tenemos

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{0.090}{0.515} = 0.175$$

Esto es menor que $p(A)$, que es 0.392. Por otro lado,

$$p(A \mid C) = \frac{p(A \cap C)}{p(C)} = \frac{0.302}{0.485} = 0.623$$

que realmente es un poco mayor que $p(A)$. Además,

$$p(B \mid A) = \frac{p(B \cap A)}{p(A)} = \frac{0.090}{0.392} = 0.23$$

$$p(C \mid A) = \frac{p(C \cap A)}{p(A)} = \frac{0.302}{0.392} = 0.77$$

En realidad, aun cuando $p(B)$ es ligeramente mayor que $p(C)$, $p(B \mid A)$ es mucho menor que $p(C \mid A)$.

❐

Ejemplo 3.31

Una moneda escogida al azar es lanzada. La probabilidad de que la moneda escogida no estuviera cargada y de que se obtenga cara es de un tercio. La probabilidad de que se escogiera una moneda no cargada y de que se obtenga cruz es también de un tercio. La probabilidad de que una moneda cargada fuese escogida y de que se obtenga cara es de un doceavo. La probabilidad de que una moneda cargada fuese escogida y de que se obtenga cruz es de un cuarto.

Es claro, la probabilidad de obtener cara es

$$\frac{1}{3}+\frac{1}{12}=\frac{5}{12}$$

y la probabilidad de que una moneda cargada fuese escogida es

$$\frac{1}{12}+\frac{1}{4}=\frac{1}{3}$$

Así, la probabilidad condicional de que una moneda cargada fuese escogida y de que se obtiene cara es

$$\frac{1/12}{5/12}=\frac{1}{5}$$

y la probabilidad condicional de que se obtenga cara dado que una moneda cargada fuese escogida es

$$\frac{1/12}{1/3}=\frac{1}{4}$$

❐

Ejemplo 332

Se tiraron tres dados. Puesto que no hubo dos caras iguales, ¿cuál es la probabilidad de que haya resultado un as? Si A denota el evento de que hubiera un as, y B el evento de que no hubo dos caras iguales, notamos que

$$p(B)=\frac{P(6,3)}{6^3} \qquad p(A\cap B)=\frac{3P(5,2)}{6^3}$$

así,

$$p(A\mid B)=\frac{3P(5,2)}{P(6,3)}=\frac{1}{2}$$

❐

Debemos señalar que, en general, $p(A|\ B)$ no es igual a $p(B \setminus A)$. Consideremos el ejemplo de tirar un dado: si A denota el evento de que el 5 aparezca, y B de que aparezca un número impar. Es obvio, $p(A\mid B)=\frac{1}{3}$, en tanto que $p(B\mid A)=1$.

*3.8 INFORMACIÓN E INFORMACIÓN MUTUA

Supongamos que fuimos informados de que se tiró un dado y el resultado fue 4. Es claro que nos dieron toda la información referente al resultado del experimento. Si nos hubieran informado que el resultado fue rojo, coincidiríamos en que nos fue dada cierta información pero no toda (el resultado se reduce a una de dos posibilidades[†]). Por otro lado, si nos

[†] Para los no apostadores, las caras 1 y 4 del dado son rojas, y las caras 2,3, 5 y 6 son negras.

informaran que el resultado fue negro, sentiríamos que también nos fue dada una información parcial, pero aún menor (el resultado se reduce a una de cuatro posibilidades).

Supongamos que después de seis semanas de iniciado el semestre se informa a los estudiantes que habrá un examen de una hora. Es claro que el anuncio contiene cierta cantidad de información. Sin embargo, si a los estudiantes se les informa después de una semana de clases que habrá un examen de una hora, podríamos decir que el anuncio contiene mucha más información debido a que es realmente inesperado que un examen pueda ser programado después de una semana de clases.

Estos ejemplos muestran que es deseable una medida cuantitativa acerca de cuánta información proporciona una fracción de un mensaje (suponemos que un mensaje es siempre una aseveración *infalible).* Si un enunciado nos informa de la ocurrencia de cierto evento que puede suceder, prodríamos decir que el enunciado contiene sólo una pequeña cantidad de información. Por otro lado, si un enunciado nos informa de la ocurrencia de cierto evento que no es probable que suceda, entonces podríamos decir que el enunciado contiene una gran cantidad de información. Lo anterior sugiere que la información contenida en un enunciado que afirma la ocurrencia de un evento, depende de la probabilidad de ocurrencia del evento. Definimos la información contenida en un enunciado que asevera la ocurrencia de un evento como

$$-\lg\ p^{\dagger}$$

donde p es la probabilidad de ocurrencia del evento. Primero observemos que como p siempre es menor o igual a 1, $\lg p$ siempre es un número no-positivo. Consecuentemente, $-\lg p$ siempre será un número no-negativo. Más aún, de inmediato se hace obvio que conforme el valor de p sea menor, mayor será la cantidad de $-\lg p$, lo cual es, en efecto, lo que queríamos.[‡]

Así, por ejemplo, cuando fuimos informados de que el resultado de tirar un dado fue 4, la cantidad de información que recibimos pudo calcularse como

$$-\lg \frac{1}{6} = \lg 6 = 2.585$$

Por otro lado, cuando fuimos informados de que el resultado fue rojo, la cantidad de información que recibimos pudo calcularse como

$$-\lg \frac{2}{6} = \lg 3 = 1.585$$

Suponga que recibimos como salida de una computadora un dígito binario, ya sea 0 o 1,

† Usamos lg para denotar el logaritmo base 2.

‡ Sin embargo, un lector podría señalar de inmediato que existen muchas otras maneras de definir una medida de información, la cual se incrementa conforme el valor de p decrece. Por ejemplo, $1/p$, $1/p^2$, y $1-p$ son sólo algunas de las muchas posibilidades que podemos escoger. Como uno confirmaría en un curso sobre teoría de información, nuestra opción es natural porque ésta tiene otras propiedades que concuerdan muy bien con nuestra intuición.

con igual probabilidad de ocurrencia. Cuando nos informan que la salida es 1, la cantidad de información que recibimos es

$$-\lg\frac{1}{2}=1$$

De modo similar, cuando nos informan que la salida es 0, la cantidad de información que recibimos es

$$-\lg\frac{1}{2}=1$$

En efecto, cuando usamos la fórmula -lg p para calcular la información contenida en un enunciado, la unidad será denominada *bit* (reducción de *bi*nary digi*t*), dado que ésta es la cantidad de información transportada por un (igualmente probable) dígito binario.

Ahora, suponga que recibimos 32 dígitos binarios como la salida de una computadora. Si se piensa que todas las 2^{32} posibilidades son igualmente probables, entonces la información que recibimos es

$$-\lg\frac{1}{2^{32}}=32 \text{ bits}$$

Nuestra discusión nos permite introducir la noción de información mutua. Suponga que nos informaron de que el resultado de tirar un dado es rojo. ¿Qué tanto nos ayuda determinar que el resultado es 4? Suponga que fuimos informados de que el profesor se encontrará fuera de la ciudad mañana. ¿Qué tanto nos ayuda determinar que habrá un examen de una hora mañana? Así, deseamos saber la cantidad de información relativa a la ocurrencia del evento A que está contenida en un enunciado que asevera la ocurrencia del evento B, la cual denotaremos por $I(A, B)$. Dado que -lg $p(A)$ es la cantidad de información contenida en un enunciado que asevera la ocurrencia de A y -lg $p(A \mid B)$ es la cantidad de información contenida en un enunciado que asevera la ocurrencia de A dado que ha ocurrido B, la diferencia entre estas dos cantidades es la cantidad de información sobre la ocurrencia de A proporcionada por una aseveración de que B ha ocurrido. En otras palabras, necesitamos -l*gp(A)* bits de información para aseverar la ocurrencia del evento A, y aún necesitaremos -*lgp(A* | *B)* bits de información para aseverar la ocurrencia del evento A después de haber sido informados de que el evento B ha ocurrido. Así, la información proporcionada por la ocurrencia del evento B sobre la ocurrencia del evento A es

$$I(A, B)=[-\lg p(A)]-[-\lg p(A \mid B)]=-\lg p(A)+\lg P(A \mid B) \tag{3.4}$$

Por ejemplo, si A es el evento en que aparece el 4 y B es el evento en que aparece un rojo cuando se ha tirado un dado, entonces

$$\begin{aligned} I(A, B) &= -\lg p(A)+\lg P(A \mid B) \\ &= -\lg\frac{1}{6}+lg\frac{1}{2} \\ &= 2.585-1 \\ &= 1.585 \text{ bits} \end{aligned}$$

Por otro lado, si C es el evento de que apareciera un número par, entonces

$$\begin{aligned} I(A, C) &= -\lg p(A) + \lg p(A \mid C) \\ &= -\lg \frac{1}{6} + \lg \frac{1}{3} \\ &= 2.585 - 1.585 \\ &= 1 \text{ bit} \end{aligned}$$

Examinemos (3.4) en forma más cuidadosa para que podamos comprender mejor la importancia de la definición de información mutua. Si $p(A \setminus B)$ es grande, esto significa que la ocurrencia de B indica una fuerte posibilidad de ocurrencia de A. En consecuencia, $I(A, B)$ es grande.* Sin embargo, si $p(A \setminus B)$ es pequeña, significa que la ocurrencia de B no dice mucho acerca de la ocurrencia de A. En consecuencia, $I(A, B)$ es pequeña. De hecho, la ocurrencia del evento B puede significar que es menos probable que el evento A ocurra. En tal caso, $p(A \mid B)$ es menor *que* $p(A)$ e $I(A, B)$ es una cantidad negativa, como veremos en el ejemplo 3.33.

Atendamos algunos casos extremos. Supongamos que B es un subconjunto de A en S. En este caso, de modo intuitivo, la ocurrencia de B asegura la ocurrencia de A. Dado que tenemos $p(A \cap B) = p(B)$, se sigue que $p(A \mid B) = 1$ y $-\lg p(A \mid B) = 0$; esto es, la información mutua proporcionada por la aseveración de que B ha ocurrido sobre la ocurrencia de A es igual a la información proporcionada por la aseveración de que A ha ocurrido. No obstante, supongamos que B es la totalidad del espacio muestral. En tal caso, $p(A \cap B) = p(A)$ y $-\lg p(A \mid B) = -\lg p(A)$. Efectivamente, $1(A, B) = 0$ significa que la ocurrencia de B no nos dice nada acerca de la ocurrencia de A.

Consideremos algunos ejemplos más:

Ejemplo 3.33

Considérese el problema de estimar la factibilidad de que haya un examen de una hora cuando el profesor está programado para salir fuera de la ciudad. Sea 5= $\{x_x, x_2, x_3, x_4\}$ el espacio muestral, donde los puntos muéstrales representan los posibles resultados:

x_1: profesor fuera de la ciudad y examen realizado
x_2: profesor fuera de la ciudad y examen no realizado
x_3: profesor en la ciudad y examen realizado
x_4: profesor en la ciudad y examen no realizado

Más aún,

$$p(x_1) = \frac{1}{2}$$

$$p(x_2) = \frac{1}{16}$$

† Recordamos al lector que lg $p(A \mid B)$ es una cantidad negativa, y que $|\lg p(A \mid B)|$ es pequeña cuando $p(A \mid B)$ es grande.

$$p(x_3) = \frac{3}{16}$$

$$p(x_4) = \frac{1}{4}$$

Si *A* denota el evento de que se realice el examen, y *B* el evento de que el profesor esté fuera de la ciudad, tenemos que

$$p(A) = \frac{1}{2} + \frac{3}{16} = \frac{11}{16}$$

$$p(A \mid B) = \frac{1/2}{1/2 + 1/16} = \frac{8}{9}$$

La información necesaria para determinar que un examen será realizado es

$$\begin{aligned} -\lg p(A) &= -\lg \frac{11}{16} \\ &= -\lg 11 + \lg 16 \\ &= -3.46 + 4 \\ &= 0.54 \text{ bits} \end{aligned}$$

y la información proporcionada por el hecho de que el profesor esté fuera de la ciudad dado el hecho de que un examen será realizado es

$$\begin{aligned} I(A, B) &= -\lg \frac{11}{16} + \lg \frac{8}{9} \\ &= -\lg 11 + \lg 16 + \lg 8 - \lg 9 \\ &= -3.46 + 4 + 3 - 3.17 \\ &= 0.37 \text{ bits} \end{aligned}$$

Si C denota el evento de que el profesor se encuentra en la ciudad, dado que

$$p(A \mid C) = \frac{3/16}{3/16 + 1/4} = \frac{3}{7}$$

tenemos

$$\begin{aligned} I(A, C) &= -\lg \frac{11}{16} + \lg \frac{3}{7} \\ &= -3.46 + 4 + 1.58 - 2.81 \\ &= -0.69 \text{ bits} \end{aligned}$$

El hecho de que el profesor se encuentre en la ciudad hace menos posible que un examen se realice. Por tanto, la información mutua proporcionada por la presencia del profesor sobre la ocurrencia del examen es una cantidad negativa. ❐

Ejemplo 3.34

La figura 3.2 muestra un modelo simple de un canal de comunicación conocido como el *canal binario simétrico*. En el extremo de transmisión, ya sea el 0 o el 1 son transmitidos, y en el extremo de recepción, ya sea el 0 o el 1 son recibidos. Específicamente, cuando un 0 es transmitido la probabilidad de que un 0 sea recibido es $1-\varepsilon$, y la probabilidad de que un 1 sea recibido es ε. Cuando un 1 es transmitido la probabilidad de que un 1 sea recibido es $1-\varepsilon$, y la probabilidad de que un 0 sea recibido es ε. Suponga que tenemos dos mensajes igualmente probables m_1 y m_2 que serán transmitidos por el canal usando las representaciones 000 y 111, respectivamente. Si 010 fue recibido, podemos calcular la información mutua entre el evento de que el mensaje m_1 fue transmitido y el evento de que parte o la totalidad de la secuencia 010 fue recibida.

$$I(m_1, 0) = -\lg\frac{1}{2} + \lg\frac{\frac{1}{2}(1-\varepsilon)}{\frac{1}{2}(1-\varepsilon)+\frac{1}{2}\varepsilon} = 1 + \lg(1-\varepsilon)$$

$$I(m_1, 01) = -\lg\frac{1}{2} + \lg\frac{\frac{1}{2}(1-\varepsilon)\varepsilon}{\frac{1}{2}\varepsilon(1-\varepsilon)+\frac{1}{2}\varepsilon(1-\varepsilon)} = 0$$

$$I(m_1, 010) = -\lg\frac{1}{2} + \lg\frac{\frac{1}{2}(1-\varepsilon)^2\varepsilon}{\frac{1}{2}\varepsilon(1-\varepsilon)^2+\frac{1}{2}\varepsilon^2(1-\varepsilon)} = 1 + \lg(1-\varepsilon)$$

El saber que 0 o 010 fue recibido nos dice exactamente la misma cantidad de información sobre la transmisión del mensaje m_1 Sin embargo, el saber que la secuencia 01 fue recibida no nos dice nada acerca de la transmisión del mensaje m_1 Intuitivamente, esto es lo que podríamos esperar, ya que la transmisión de m_1 o m_2 podría dar origen a la secuencia 01 en el extremo de la recepción con la misma probabilidad. ❐

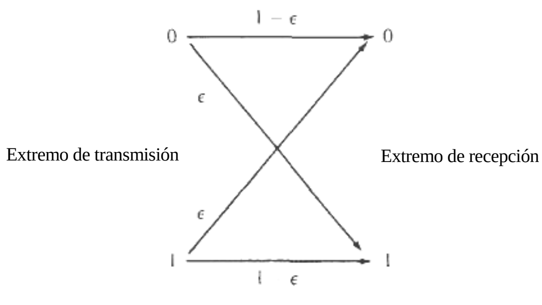


Figura 3.2

Al observar que

$$
\begin{aligned}
I(A, B) &= -\lg p(A) + \lg p(A \mid B) \\
&= -\lg p(A) - \lg p(B) + \lg p(A \cap B) \\
&= -\lg p(B) + \frac{\lg p(A \cap B)}{\lg p(A)} \\
&= -\lg p(B) + \lg p(B \mid A) \\
&= I(B, A)
\end{aligned}
$$

Nos damos cuenta de que la información mutua es una medida simétrica sobre la información relativa a dos eventos. En otras palabras, lo que la ocurrencia de *B* nos dice acerca de la ocurrencia de *A* es igual a lo que la ocurrencia de *A* nos dice acerca de la ocurrencia de *B*. Así, *I(A, B)* es una medida de la *información mutua* de *B* hacia *A* como también de *A* hacia *B*.

3.9 NOTAS Y REFERENCIAS

Para una referencia general sobre combinatoria, véase Berge [1], Bermany Fryer [2], Bogart [3], Cohén [4], Even [5], Liu [9], Riordan [11], Ryser [12], Tucker [13] y Vilenkin [14]. Consulte en Whitworth [15,16] algunas extensiones y más ejemplos en el área de permutaciones y combinaciones. Consulte los capítulos 1 y 2 de Even [5], y el capítulo 5 de Reingold, Nievergelt, y Deo [10] sobre algoritmos para generar permutaciones y combinaciones de un conjunto dado de objetos. Feller [7] es una excelente referencia para teoría de probabilidad. Para el tópico de teoría de la información, véase Fano [6] y a Gallager [8].

1. Berge, C.: *Principles of Combinatorios,* Academic Press, Nueva York, 1971.
2. Berman, G. y K. D. Fryer: *Introduction to Combinatorios,* Academic Press, Nueva York, 1972.
3. Bogart, K. P: *Introductory Combinatorios,* Pitman Publishing, Marshfield, Mass., 1983.
4. Cohén, I. A. C: *Basic Techniques of Combinatorial Theory,* John Wiley & Sons, Nueva York, 1978.
5. Even, S.: *Algorithmic Combinatorios,* Macmillan Company, Nueva York, 1973.
6. Fano, R. M.: *Transmission of Information,* M.I.T. Press, Cambridge, Mass., 1961.
7. Feller, W.: *An Introduction to Probability Theory andlts Applications, 2"* ed., John Wiley & Sons, Nueva York, 1950.
8. Gallager, R. G.: *Information Theory and Reliable Communication,* John Wiley & Sons, Nueva York, 1968.
9. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
10. Reingold, E. M., J. Nievergelt y N. Deo: *Combinatorial Algorithms: Theory and Practice,* Prentice-Hall, Englewood Cliffs, N. J., 1977.
11. Riordan, J.: *An Introduction to Combinatorial Analysis,* John Wiley & Sons, Nueva York, 1958.
12. *Ryser,H.J.: Combinatorial Mathematics,* publicado por laMathematical Associationof America, distribuido por John Wiley & Sons, Nueva York, 1963.

13. Tucker, A.: *Applied Combinatorics,* John Wiley & Sons, Nueva York, 1980.
14. Vilenkin, N. Ya.: *Combinatorios* (traducido del ruso por A. Shenitzer y S. Shenitzer), Academic Press, Nueva York, 1971.
15. Whitworth, W. A.: *Choice and Chance,* reimpreso de la 5ª ed. (1901), Hafner Publishing Company, Nueva York, 1965.
16. Whitworth, W. A.: *DCC Exercises in Choice and Chance*, reimpreso de la ed. 1897, Hafher Publishing Company, Nueva York, 1965.

PROBLEMAS

3.1 Un menú en un restaurante se lee como sigue:

Grupo A: Sopa Wonton
Sopa de aleta de tiburón
Rollos de primavera
Rumayki

Grupo B: Pato almendrado
Chow Mein de pollo
Moo Goo Gai Pan

Grupo C: Cerdo agridulce
Bistec a la pimienta
Res a la dragón
Camarones mariposa
Camarones en salsa de langostino
Egg Foo Young

Grupro D: Café
Té
Leche

a) Suponga que selecciona una opción de cada grupo sin omisión o sustitución. ¿Cuántas' 'cenas completas de 4 platillos" diferentes pueden formarse a partir de este menú?

b) Suponga que el mesero no lo fuerza a realizar una selección si desea omitir completamente un grupo (después de todo, usted está pagando). ¿Cuántas cenas diferentes puede formar a partir de este menú?

c) Suponga que selecciona una opción de cada uno de los grupos *A, B* y *D,* y dos opciones del grupo *C* sin omisiones o sustituciones. ¿Cuántas cenas diferentes puede idear? Suponga que selecciona una o dos opciones del grupo *C* sin cualquier otra omisión o remplazo. ¿Cuántas cenas puede idear?

3.2 *a)* ¿De cuántas maneras pueden seleccionarse dos enteros de entre los enteros 1, 2,..., 100 de manera que su diferencia sea exactamente siete? *b)* Repita el inciso *a)* si la diferencia debe ser menor o igual a siete.

3.3 ¿De cuántas maneras pueden seleccionarse dos cuadros adyacentes de un tablero de ajedrez de 8×8?

3.4 El nombre de una variable en un lenguaje de programación dado debe ser una letra o bien una letra seguida de un dígito decimal. ¿Cuántos nombres de variables diferentes hay en este lenguaje?

3.5 En una fila se encuentran sentados cinco niños y cinco niñas. Calcule de cuántas maneras podrán estar sentados si:

a) Todos lo niños deben estar sentados en los cinco asientos del extremo izquierdo.
b) No puede estar sentado un niño junto a otro.
c) Juan y María deben estar sentados uno junto al otro.

3.6 *a)* ¿De cuántas maneras pueden formar una fila 10 niños y 5 niñas de modo que no haya dos niñas una junto a la otra? (Todos los niños y niñas son distintos.) *b)* Repita el inciso *a)* si forman un círculo.

3.7 ¿De cuántas maneras pueden ser arregladas las letras del alfabeto inglés de manera que haya exactamente siete letras entre las letras *a* y 6?

3.8 *a)* ¿De cuántas maneras pueden permutarse las letras *a, a, a, a, a, b, c, d, e* de forma tal que no haya dos letras *a* adyacentes?

b) Repita el inciso *a)* considerando que no pueden estar adyacentes dos de las letras *b, c, d, e.*

3.9 *a)* ¿De cuántas maneras pueden acomodarse las letras de la palabra *MISSISSIPPI?*
b) ¿De cuántas maneras pueden arreglarse las letras si las dos letras *P* deben estar separadas?

3.10 *a)* ¿De cuántas maneras pueden ser acomodadas las letras *a, b, c, d,* e,/de forma que la letra *b* siempre se encuentre a la izquierda de la letra *el tí)* Repita el inciso *a)* si la letra *b* siempre se encuentra a la izquierda de la letra *e.*

3.11 *a)* Suponga que no se permiten las repeticiones. ¿Cuántos números de cuatro dígitos pueden formarse a partir de los seis dígitos 1, 2, 3, 5, 7 y 8?
b) ¿Cuántos números del inciso *á)* son menores que 4000?
c) ¿Cuántos números del inciso *a)* son pares?
d) ¿Cuántos números del inciso *a)* son impares?
e) ¿Cuántos números del inciso *á)* son múltiplos de 5?
f) ¿Cuántos números del inciso *a)* contienen tanto al dígito 3 como al dígito 5?

3.12 Un examen contiene 15 preguntas de "verdadero o falso". ¿De cuántas maneras diferentes puede un estudiante realizar el examen si se le permite no contestar algunas de las preguntas?

3.13 Un palíndromo es una palabra que se lee igual de izquierda a derecha que al revés. ¿Cuántos palíndromos de siete letras pueden construirse a partir del alfabeto inglés?

3.14 *a)* ¿Cuántas placas de automóvil pueden formarse con dos letras seguidas de cuatro dígitos?
b) Repita el inciso *a)* si las dos letras deben ser distintas.

3.15 *a)* ¿De cuántas maneras se puede formar el patrón

$$\begin{array}{ccccc} & & \times & & \\ \times & \times & \times & \times & \times \\ & & \times & & \\ & & \times & & \\ & & \times & & \end{array}$$

usando ceros y unos?
b) ¿Cuántos de estos patrones no son simétricos con respecto al eje vertical?

3.16 *a)* A partir de una baraja de 52 cartas se sacan cartas sin regresarlas. ¿De cuántas maneras pueden sacarse 10 cartas para que la décima carta sea la primera repetición?
b) Repita el inciso *a)* siendo la décima carta una repetición.

3.17 En una fila de 20 asientos, ¿de cuántas maneras se pueden formar tres bloques de asientos con secutivos con cinco asientos en cada bloque?

3.18 *a)* Mostrar que el número total de permutaciones de *p* pelotas rojas y 0, o 1, o 2, ... ,*o q* pelotas blancas es

$$\frac{p!}{p!} + \frac{(p+1)!}{p!1!} + \frac{(p+2)!}{p!2!} + \cdots + \frac{(p+q)!}{p!q!}$$

b) Mostrar que la suma en el inciso *a)* es

$$\frac{(p+q+1)!}{(p+1)!\,q!}$$

c) Mostrar que el número total de permutaciones de 0, o 1, o 2,..., o *p* pelotas rojas con 0, o 1, o 2,.. ., o *q* pelotas blancas es

$$\frac{(p+q+2)!}{(p+1)!\,(q+1)!} - 2$$

3.19 En una clase de 100 estudiantes, 40 son varones.
a) ¿De cuántas maneras se puede formar un comité de 10 personas?
b) Repita el inciso *a)* si debe haber un número igual de mujeres y varones en el comité.
c) Repita el inciso *a)* si el comité debe estar constituidora *sea* por seis varones y cuatro mujeres *o* por cuatro varones y seis mujeres.

3.20 Un estudiante debe contestar 8 de 10 preguntas en un examen.

a) ¿Cuántas opciones tiene el estudiante?

b) ¿Cuántas opciones tiene si debe contestar las primeras tres preguntas?

c) ¿Cuántas opciones tiene si debe contestar al menos cuatro de las primeras cinco preguntas?

3.21 Una delegación de cuatro estudiantes ha sido seleccionada de entre un total de 12 estudiantes para asistir a un congreso.

a) ¿De cuántas maneras puede escogerse la delegación?

b) Repita el inciso *a)* considerando que hay dos estudiantes que se niegan a estar en la delegación simultáneamente.

c) Repita el inciso *a)* considerando que hay dos estudiantes que asistirían al congreso solamente si van juntos.

d) Repita el inciso *a)* considerando que hay dos estudiantes que se niegan a estar juntos en la delegación y otros dos estudiantes que asistirían al congreso sólo si van juntos.

3.22 *a)* De entre 200 automóviles, 30 son seleccionados para probar si cumplen o no las especificaciones de seguridad. Además, 30 (de entre los mismos 200 automóviles) son seleccionados para probar si cumplen o no las especificaciones de anticontaminación. ¿De cuántas maneras puede llevarse a cabo la selección?

b) ¿De cuántas maneras se puede llevar a cabo la selección para que sólo cinco automóviles sean sometidos a ambas pruebas?

3.23 Un hombre tiene 10 amigos. ¿De cuántas maneras puede salir a cenar con dos o más de ellos?

3.24 *a)* Quince jugadores de baloncesto serán seleccionados por los tres equipos profesionales de Boston, Chicago y Nueva York; cada equipo seleccionará a cinco jugadores. ¿De cuántas maneras se puede realizar lo anterior?

b) Quince jugadores de baloncesto serán divididos en tres equipos de cinco jugadores cada uno. ¿De cuántas maneras puede realizarse esto?

3.25 ¿De cuántas maneras podemos distribuir 15 libros diferentes entre Pilar, Carolina y Luis para que Pilar y Carolina juntas reciban el doble de libros que Luis?

3.26 De entre todos los números decimales de siete dígitos, ¿cuántos de ellos contienen exactamente tres nueves?

3.27 *a)* ¿De cuántas maneras pueden seleccionarse dos números de entre los enteros 1,2, . . ., 100 de forma que su suma sea un número par?, ¿un número impar?

b) Use argumentos combinatorios para mostrar que

$$C(2n, 2) = 2C(n, 2) + n^2$$

3.28 Hay 50 estudiantes tanto de tercer grado como de cuarto grado. En cada grado hay 25 varones y 25 mujeres. ¿De cuántas maneras pueden ser seleccionados ocho representantes para que haya cuatro mujeres y tres de tercer grado?

3.29 Se seleccionan tres enteros de entre 1,2, . . ., 1000. ¿De cuántas maneras pueden seleccionarse estos enteros de forma que su suma sea divisible por 4?

3.30 ¿De cuántas maneras puede ser dividido en comités un grupo de ocho personas, sujetándose a la restricción de que cada persona debe pertenecer exactamente a un comité, y cada comité debe contener al menos dos personas? (Una división en comités de tres, tres, y dos personas es considerada igual que una división en comités de dos, tres, y tres personas).

3.31 Se seleccionan dos grupos, cada uno con 10 estudiantes a partir de 100 estudiantes. ¿De cuántas maneras puede realizarse la selección para que el estudiante más alto del primer grupo sea más bajo que el estudiante más bajo del segundo grupo? (Suponga que los 100 estudiantes son todos de diferentes estaturas.)

3.32 ¿Cuántos números decimales de n-dígitos tienen sus dígitos en orden decreciente? (El primer dígito de un número de n-dígitos no debe ser 0.)

3.33 ¿De cuántas maneras pueden otorgarse 22 libros diferentes a 5 estudiantes de modo que 2 de ellos tengan 5 libros y los otros 3 tengan 4 libros?

3.34 ¿De cuántas maneras pueden ser divididas $2n$ personas en n parejas?

3.35 *a)* ¿De cuántas maneras pueden seleccionarse dos cuadros de un tablero de ajedrez de 8 x 8 de forma que no se encuentren en la misma columna o en la misma fila?

b) ¿De cuántas maneras pueden seleccionarse cuatro cuadros de un tablero de ajedrez de 8 x 8 no encontrándose todos en la misma fila o columna, para formar un rectángulo?

3.36 Muestre que el producto de los k enteros sucesivos es divisible por $k!$ *(Sugerencia:* Considere el número de maneras de seleccionar k objetos de $n + k$ objetos.)

3.37 *a)* Muestre que

$$C(2n + 2, n + 1) = C(2n, n+l) + 2C(2n, n) + C(2n, n-l)$$

b) Escriba un argumento combinatorio de la igualdad en el inciso *a).*

3.38 *a)* De entre $2n$ objetos, n de ellos son idénticos. Encuentre el número de maneras de seleccionar n objetos a partir de estos $2n$ objetos.

b) De entre $3n + 1$ objetos, n de ellos son idénticos. Encuentre el número de maneras de seleccionar n objetos de estos $3n + 1$ objetos.

3.39 ¿De cuántas maneras pueden seleccionarse dos enteros de entre 1, 2,. . . , «- 1 de modo que su suma sea mayor que n?

3.40 A partir de un gran número de pesos, pesetas, soles y bolívares, ¿de cuántas maneras se pueden seleccionar cinco monedas?

3.41 *a)* Mostrar que el número de maneras de acomodar r pelotas indistinguibles en n cajas, $n \leqq r$, sin dejar alguna caja vacía es $C(r —1, n- 1)$. (Una caja puede contener un número arbitrario de pelotas.)

b) ¿De cuántas maneras puede programar un estudiante 15 horas de estudio en un periodo de 5 días, tales que estudie al menos una hora diaria?

c) ¿De cuántas maneras pueden colocarse r pelotas indistinguibles en n cajas con cada caja conteniendo al menos q pelotas?

d) Repita el inciso *b)* si el estudiante debe dedicar al menos dos horas diarias al estudio.

3.42 ¿De cuántas maneras pueden distribuirse r pelotas rojas y w pelotas blancas en n cajas de forma que cada caja contenga al menos una pelota de cada color?

3.43 ¿De cuántas maneras podemos distribuir $2t$ canicas entre cuatro cajas distintas *A,B,Cy D* para que las cajas *A y B* contengan a lo más / canicas, en tanto que las cajas C y *D* pueden contener cualquier número de canicas?

3.44 *a)* ¿Cuántas secuencias de m ceros y n unos hay?

b) ¿Cuántas secuencias existen de forma que cada 1 esté separado por al menos dos ceros? [Suponga, para este inciso, que $m \geqq 2(n - 1)$.]

3.45 Suponga que «juegos diferentes serán distribuidos entre n niños. ¿De cuántas maneras se puede hacer para que exactamente un niño no reciba juego alguno?

3.46 Nos proporcionan una caja roja, una caja azul y una caja verde. Además, se nos proporcionan 10 pelotas rojas, 10 pelotas azules y 10 pelotas verdes. Las pelotas del mismo color se consideran idénticas. Considere las siguientes restricciones:

1. Ninguna caja contiene una pelota que tenga el mismo color de la caja.
2. Ninguna caja está vacía.

Determine el número de maneras en las cuales podamos colocar las 30 pelotas dentro de las cajas de forma que:

a) Ninguna restricción ha de satisfacerse; esto es, cualquier combinación es permitida.

b) La restricción 1 ha de satisfacerse.

c) La restricción 2 ha de satisfacerse.

d) Las restricciones 1 y 2 han de satisfacerse.

3.47 *a)* r pelotas distintas serán colocadas en n cajas distintas, con las pelotas en cada caja arregladas en orden. Mostrar que hay

$$(n+r-1)(n+r-2)\cdots(n+1)n$$

maneras de hacer lo anterior.

b) ¿De cuántas maneras pueden arreglarse las letras *a, b, c, d, e, f, g, h* para que *a* se encuentre a la izquierda de *b* y *b* se encuentre a la izquierda de c?

3.48 Encuentre el número de permutaciones de las letras *a, b, c, d,e,f,g* de forma que ninguno de los patrones *beg* y *cad* aparezcan.

3.49 ¿Cuántas permutaciones hay de 10 dígitos 0,1, 2,..., 9 en las cuales el primer dígito es mayor que 1 y el último dígito es menor que 8?

3.50 ¿Cuántas permutaciones existen de las 26 letras *a,b,c,...,x,y,z* en las cuales la primera letra no es *a, b,* o c y la última letra no es *w, x, y* o z?

3.51 Dado el conjunto de letras *{a, b, c, . . . , z}* ¿cuántos 12-subconjuntos (esto es, conjuntos que tienen 12 elementos) hay que no contengan a alguno de los conjuntos *[h, o, n, w, a, i}, {r, o, n, a) y* {/, *i, u}* como subconjuntos?

3.52 Sobre una repisa fueron acomodados en orden alfabético diez libros, según el nombre del autor. ¿De cuántas maneras podrá un mono reacomodar los libros de modo que ningún libro quede colocado en su lugar original?

3.53 Entre todos los números de «-dígitos, ¿cuántos de ellos contienen los dígitos 2 y 7 pero no los dígitos 0, 1,8,9?

3.54 Si escribimos todos los números decimales de 1 a un millón, ¿cuántas veces habremos escrito el dígito 9?

3.55 Los pasos 2 y 3 en el procedimiento para generar todas las permutaciones del conjunto en la sección 3.5 pueden llevarse a cabo como sigue:

a) Examine la permutación $a_1 a_2 \ldots a_n$ elemento por elemento de derecha a izquierda. Sea a_m el elemento más a la derecha tal que $a_m < a_{m-1}$.

b) Examine la permutación elemento por elemento de derecha a izquierda nuevamente. Sea a_p el elemento más a la derecha tal que $a_m < a_p$.

c) Intercambie a_m y a_p.

d) Intercambie a_{m+1} y a_n, a_{m+2} y a_{n-1}, a_{m+3} y a_{n-2}, y así sucesivamente. [Note que después del paso *c)* el a_m original toma el lugar de a_p .]

Convénzase usted mismo de que estos pasos efectivamente dan lugar a la siguiente permutación en el orden lexicográfico.

3.56 Siete (distintos) accidentes automovilísticos ocurren en una semana. ¿Cuál es la probabilidad de que todos ellos ocurran el mismo día?

3.57 Diez personas (distintas) entran a un elevador en la planta baja de un edificio de 20 pisos. ¿Cuál es la probabilidad de que todas ellas salgan del elevador en diferentes pisos?

3.58 Cuando McGraw-Hill Book Company decidió ordenar una reimpresión de este libro, los editores fueron informados por el autor de que existían 50 errores de impresión, que deberían ser corregidos (en realidad, hubo muchos más errores que aquéllos decubiertos por el autor. No obstante, éste es otro asunto). Entre las 422 páginas en el libro, ¿cuál es la probabilidad de que estos 50 errores aparezcan en 10 o menos páginas?

3.59 A partir de los números 1,2,.. ., 100, se escoge un primer número y luego otro a partir de los números restantes. Suponiendo que las 9900 posibilidades son todas igualmente probables, ¿cuál es la probabilidad de que la suma de los dos números sea divisible por 3?

3.60 Una de 10 llaves abre una puerta. Si intentamos abrir con una llave tras otra, ¿cuál es la probabilidad de que la puerta sea abierta en el primer intento?, ¿en el segundo intento?, ¿en el décimo intento?

3.61 Hay 10 lugares de estacionamiento adyacentes en un lote de estacionamiento. Cuando usted llega en su Rolls Royce nuevo ya se encuentran siete automóviles en el estacionamiento. ¿Cuál es la probabilidad de que usted pueda encontrar dos espacios adyacentes desocupados para su Rolls Royce?

3.62 Un número es escogido al azar de entre los 30 números {10,11,… ,19,20,21,..., 30, 31,..., 39}. Se sabe que los números con el mismo primer dígito tienen igual probabilidad de ser escogidos. Además, un número con el 2 como primer dígito es doblemente probable que sea escogido que uno con 1 como el primer dígito, y un número con 3 como primer dígito es tres veces más probable que sea escogido que uno con 1 como su primer dígito.

a) Describa el espacio muestral.
b) ¿Cuál es la probabilidad de que un número cuyo primer dígito es 2 sea escogido?
c) ¿Cuál es la probabilidad de que un número cuyo segundo dígito es 2 sea escogido?
d) ¿Cuál es la probabilidad de que sea escogido un número cuyos primero o segundo dígitos, o ambos dígitos, sean 2?
e) ¿Cuál es la probabilidad de que un número divisible por 3 sea escogido?
f) ¿Cuál es la probabilidad de que un número divisible por 5 sea escogido?
g) ¿Cuál es la probabilidad de que un número divisible por 3 o por 5, pero no por ambos, sea escogido?

3.63 Una compañía aseguradora está interesada en la distribución de edad de las parejas casadas cuando la edad de ambos, esposo y esposa, se encuentra entre los 21 y 25 años. Considere el espacio muestral de los 25 puntos muéstrales, cada uno de los cuales está representado por una pareja ordenada de números *(x, y)*, donde *x* es la edad del esposo y y es la edad de la esposa:

$$S = \{(x, y) \mid 21 \leqq x \leqq 25, 21 \leqq y \leqq 25\}$$

La probabilidad asociada con el punto muestral *(x, y)* es igual a kx/y si $x \leqq y$, y a ky/x si $y \leqq x$.

a) Determine el valor de *k*.
b) Suponga que una pareja de este grupo de edades es seleccionada al azar, ¿cuál es la probabilidad de que el esposo tenga 21 años y la esposa 22 años?, ¿de que el esposo y la esposa tengan la misma edad?, ¿de que el esposo sea mayor que la esposa?, ¿de que el esposo tenga 25 años?
c) Suponga que una pareja de este grupo de edades es seleccionada aleatoriamente. ¿Cuál es la probabilidad de que al menos uno de ellos tenga 22 años o más, o de que tanto el esposo como la esposa tengan la misma edad?
d) Suponga que una pareja de este grupo de edades es seleccionada aleatoriamente. Dado que el esposo tiene 23 años, ¿cuál es la probabilidad de que la esposa tenga 24 años?, ¿cuál es la probabilidad de que la esposa sea mayor que el esposo?
e) Suponga que una pareja de este grupo de edades es seleccionada al azar. Dado que ambos, esposo y esposa, son al menos de 22 años, ¿cuál es la probabilidad de que ambos sean al menos de 24 años?, ¿de que uno o ambos sean al menos de 24 años?
f) Suponga que una pareja de este grupo de edades es seleccionada al azar. Dado que al menos uno de ellos tiene 23 años o menos, ¿cuál es la probabilidad de que ambos sean de 22 años o menos?

3.64 Considere el experimento de lanzar una moneda no cargada hasta que dos caras o dos cruces aparezcan una tras otra.

a) Describa el espacio muestral.
b) ¿Cuál es la probabilidad de que el experimento termine antes del sexto lanzamiento?
c) ¿Cuál es la probabilidad de que el experimento termine después de un número par de lanzamientos?

d) Dado que el experimento termina con dos caras, ¿cuál es la probabilidad de que el experimento termine antes del sexto lanzamiento?

e) Dado que el experimento no termina antes del tercer lanzamiento, ¿cuál es la probabilidad de que el experimento no termine después del sexto lanzamiento?

3.65 Hay 10 pares de zapatos en un clóset. Si ocho zapatos se escogen al azar, ¿cuál es la probabilidad de que no se escoja ningún par completo de zapatos?, ¿de que sea escogido exactamente un par completo de zapatos?

3.66 Hay un 30 por ciento de posibilidad de lluvia en un día particular cualquiera. ¿Cuál es la probabilidad de que haya al menos un día lluvioso dentro de un periodo de siete días? Dado que hay al menos un día lluvioso, ¿cuál es la probabilidad de que haya al menos dos días lluviosos?

3.67 Una compañía compró 100 000 transistores (50 000 del proveedor *A*, 30 000 del proveedor *B* y 20 000 del proveedor Q. Se sabe que el 2 por ciento de los transistores del proveedor *A* son defectuosos, el 3 por ciento de los transistores del proveedor *B* son defectuosos y el 5 por ciento de los transistores del proveedor *C* son defectuosos.

a) Si un transistor de los 100 000 transistores se escoge al azar, ¿cuál es la probabilidad de que sea defectuoso?

b) Dado que un transistor seleccionado aleatoriamente es defectuoso, ¿cuál es la probabilidad de que provenga del proveedor *A*?

c) Puesto que un transistor seleccionado al azar no es del proveedor *A*, ¿cuál es la probabilidad de que sea defectuoso?

3.68 Un sistema computacional está constituido de seis subsistemas. Cada subsistema puede fallar independientemente con una probabilidad de 0.2. La falla de cualquier subsistema causará la falla de todo el sistema computacional. Dado que el sistema computacional falla, ¿cuál es la proba bilidad de que sólo el subsistema 1 falle?

3.69 Hay un radar, una computadora y un giroscopio a bordo de un aeroplano. La probabilidad de que el radar falle es de 0.2. Si el radar falla, el giroscopio también fallará y la probabilidad de que la computadora falle es de 0.3. Si el radar funciona adecuadamente entonces la computadora también funcionará correctamente y la probabilidad de que el giroscopio falle es 0.2.

a) Describa el espacio muestral.

b) ¿Cuál es la probabilidad de que la computadora o el giroscopio funcionen correctamente en tanto que el otro no lo hace?

c) ¿Cuál es la probabilidad de que el radar funcione en forma correcta si uno de los otros dos sistemas falla?

3.70 Considere el problema de seleccionar un número de entre los 30 números descritos en el problema 3.62. Determine la información en cada una de las aseveraciones en *a), b), c)* y *d).*

a) El número 27 fue escogido.

b) Un número con 1 como su primer dígito fue escogido.

c) Un número entre 25 y 30 (inclusive) fue escogido.

d) Un número de dos dígitos, la suma de cuyos dígitos es 9, fue escogido.

e) ¿Cuál es la información mutua entre los dos eventos en *a) y b)?,* ¿en *a)* y *c)*?, ¿en *a)* y *d)?* Calcule la información mutua de dos maneras diferentes.

3.71 Considere el problema sobre la distribución de edades de parejas casadas en el problema 3.63. Determine la información en cada una de las aseveraciones en *a), b)* y c).

a) El esposo es mayor que la esposa.

b) La diferencia entre las edades del esposo y la esposa es menor o igual a 2.

c) El esposo tiene 25 arlos o la esposa tiene al menos 22 años, o ambas.

d) ¿Cuál es la información mutua entre los dos eventos en *a)* y *b*)?, ¿en *a)* y c)?, ¿en *b)* y *c*)? Calcule la información mutua de dos maneras diferentes.

CAPÍTULO

CUATRO

Relaciones y funciones

4.1 INTRODUCCIÓN

En muchos problemas referentes a objetos discretos, con frecuencia se da el caso de que existe algún tipo de relación entre los objetos. En un conjunto de programas de computadora, podríamos decir que dos programas están relacionados si ambos utilizan algunos datos en común, y si no es así, no están relacionados. Entre un grupo de estudiantes podríamos decir que dos estudiantes están relacionados, si las primeras letras de sus apellidos son las mismas, o por lo contrario, podríamos establecer que dos estudiantes están relacionados si las primeras letras de sus apellidos son diferentes. De igual manera, al considerar el conjunto de enteros {1,2,3,...,15}. Podríamos decir que tres enteros en el conjunto están relacionados si su suma es divisible por 5. Así, los enteros 2, 3, 5 están relacionados y los enteros 5, 10, 15 están relacionados, pero los enteros 1, 2, 4 no lo están. En este capítulo estudiamos relaciones entre objetos discretos.

En el capítulo 2 introdujimos la noción de un par ordenado de objetos. Sean *A y B* dos conjuntos. El *producto cartesiano* de *A* y *B*, denotado por *A* x *B*, es el conjunto de todos los pares ordenados de la forma *(a, b)* donde $a \in A$ y $b \in B$. Por ejemplo,

$$\{a, b\} \times \{a, c, d\} = \{(a, a), (a, c), (a, d), (b, a), (b, c), (b, d)\}$$

Una *relación binaria* de *A* sobre *B* es un subconjunto de *A x B.* Una relación binaria es en realidad sólo una formalización de la noción intuitiva de que algunos elementos de *A* están relacionados con algunos elementos de *B.* De hecho, si *R* es una relación binaria de *A* a *B y* si el par ordenado *(a, b)* está en *R,* diremos que el elemento *a está elacionado con* el elemento *b.* Por ejemplo, sean *A* = *{a,b,c,d}* un conjunto de cuatro estudiantes, y 5= {*CS*121,*GS*221, *CS*257, *CS*264, *CS*273, *CS*281} un conjunto de seis cursos. El producto cartesiano *A x B* contiene todas las posibles parejas de estudiantes y cursos. Por otro lado, una relación

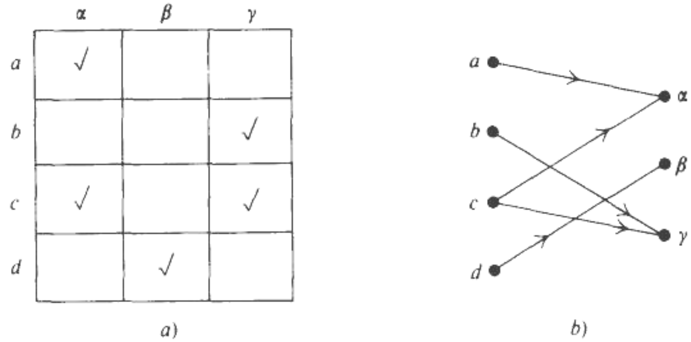


Figura 4.1

R = {(*a*, *CS*121), (*b*, *CS*221), {*b*, *CS*264), (*c*, *CS*221), (*c*, *CS*257), (*c*, *CS*273), (*d*, *CS*257), (*d*, *CS*281)} podría describir los cursos que los estudiantes llevan, y la relación *T*= {(*a*, GS121), (c, *CS*257), (c, *CS*273)} podría describir los cursos con los cuales los estudiantes tienen dificultades.

Junto a una lista de pares ordenados, una relación binaria también puede representarse en forma tabular o gráfica. Por ejemplo, sean $A = \{a, b, c, d\}$ y $B = \{\alpha, \beta, \gamma\}$, y si $R = \{(a, \alpha), (b, \gamma), (c, \alpha), (c, \gamma), (d, \beta)\}$ una relación binaria de *A* a *B*. *R* puede representarse en forma tabular, como se muestra en la figura 4.1*a*, donde las filas de la tabla corresponden a los elementos en *A*, *y* las columnas de la tabla corresponden a los elementos en *B;* una marca en una celda significa que el elemento de la fila que contiene la celda está relacionado con el elemento de la columna que contiene la celda. *R* también puede representarse en forma gráfica como se muestra en la figura 4.1 *b*, donde los puntos de la columna del lado izquierdo son los elementos de *A*, los puntos en la columna del lado derecho son los elementos de *B*, y una flecha desde un punto en la columna del lado izquierdo hacia un punto en la columna del lado derecho indica que el elemento correspondiente de *A* está relacionado con el elemento correspondiente de *B*.

Puesto que las relaciones binarias son conjuntos de pares ordenados, las nociones de intersección de dos relaciones, unión de dos relaciones, diferencia simétrica de dos relaciones y diferencia de dos relaciones se obtienen directamente a partir de las correspondientes para conjuntos. Para ser específicos, sean R_1 y R_2 dos relaciones binarias de *A* a *B*. Entonces $R_1 \cap R_2$, $R_1 \cup R_2$ $R_1 \oplus R_2$ y $R_1 - R_2$ también son relaciones binarias de *A* a *B*, las cuales se conocen como la intersección, unión, diferencia simétrica y diferencia de R_1 *y* R_2. Por ejemplo, *sea A* = {a, *b*, *c*, *d]* un conjunto de estudiantes y *B* = {*CS*121, *CS*221, *CS*257, *CS*264, *CS*273, *CS*281} un conjunto de cursos. Podríamos tener una relación binaria R_1, de *A* a *B*, al describir los cursos que los estudiantes llevan, y una relación binaria R_2 de *A* a *B* que describa los cursos en los cuales los estudiantes están interesados, como se muestra en la figura 4.2. Entonces, la relación binaria $R_1 \cap R_2$, la cual es {{*a*, *CS*121), (*b*, *CS*221), {*d*, *CS*264), (*d*, *CS*281)}, describe los cursos que los estudiantes llevan y en los cuales, además, están interesados. La relación binaria $R_1 \cup R_2$, que es {(*a*, *CS*121), (*a*, *CS*264), (*b*, *CS*221), (*b*, *CS*257), (*b*, *CS*273), (*c*, *CS*221), (*c*, *CS*273), (*c*, *CS*281), {*d*, *CS*264), (*d*, *CS*273), (*d*,

	$CS121$	$CS221$	$CS257$	$CS264$	$CS273$	$CS281$
a	✓					
b		✓	✓			
c		✓			✓	✓
d				✓		✓

R_1

	$CS121$	$CS221$	$CS257$	$CS264$	$CS273$	$CS281$
a	✓			✓		
b		✓			✓	
c						
d				✓	✓	✓

R_2

Figura 4.2

*CS*281)}, describe los cursos que los estudiantes llevan o en los cuales están interesados. La relación binaria $R_1 \oplus R_2$, que es *{(a, CS264), (b, CS251), (b, CS*273), (*c*, *CS*221), *(c*, *CS*273), *(c*, *CS*281), *(d*, *CS*273)}, describe los cursos en los cuales los estudiantes están interesados pero no los llevan o los llevan pero no están interesados en ellos. La relación binaria R_x - R_2, la cual es {(*b*, *CS*257), (*c*, *CS*221), (*c*, *CS*273), (*c*, *CS*281)}, describe los cursos que los estudiantes llevan pero en los cuales no están interesados.

Otro ejemplo: sea *A* = *{a, b, c, d)* un conjunto de estudiantes y *B* = {BT&T, CompComm, GEE,JBM, Orange} un conjunto de compañías que acuden a la universidad con el fin de entrevistar a estudiantes para empleos. Podríamos tener una relación binaria R_1 de *A* a *B* al especificar las entrevistas que las compañías tuvieron con los estudiantes, y una relación binaria R_2 de *A* a *B* especificando las ofertas de trabajo que las compañías hicieron a los estudiantes, como se muestra en la figura 4.3. Invitamos al lector para que encuentre el significado de las relaciones binarias $R_1 \cap R_2$, $R_1 \cup R_2$, $R_1 \oplus R_2$ y $R_1 - R_2$.

Como las relaciones binarias describen la relación entre pares de objetos, nos gustaría definir las relaciones ternarias para describir la relación entre tripletas de objetos, y relaciones cuaternarias para describir la relación entre cuádruplas de objetos, y así sucesivamente. De esta manera, una *relación ternaria* entre tres conjuntos *A, B y C* está definida como un subconjunto del producto cartesiano de los dos conjuntos *A* x*B y C,* denotado por *(A* x *B)* x C. Observemos que *(A x B)* x *C* es el conjunto de todas las tripletas ordenadas de la forma $((a, b), c)$, donde $(a, b) \in A \times B$ y $c \in C$. Por ejemplo, si $A = \{a, b\}$, $B = \{\alpha, \beta\}$ y $C = \{1, 2\}$. Tenemos

$$
\begin{aligned}
(A \times B) \times C = \{&((a, \alpha), 1), ((a, \alpha), 2), ((a, \beta), 1), ((a, \beta), 2), \\
&((b, \alpha), 1), ((b, \alpha), 2), ((b, \beta), 1), ((b, \beta), 2)\}
\end{aligned}
$$

	BT&T	Compcomm	GEE	JBM	Orange
a	✓	✓		✓	
b	✓	✓	✓	✓	✓
c				✓	✓
d	✓		✓	✓	

R_1

	BT&T	Compcomm	GEE	JBM	Orange
a					
b			✓	✓	
c		✓		✓	✓
d			✓	✓	✓

R_2

Figura 4.3

Sean A un conjunto de estudiantes, B un conjunto de cursos y C el conjunto de todas las calificaciones posibles. Entonces una relación ternaria entre A,*ByC* puede definirse para especificar las calificaciones que los estudiantes obtuvieron en los cursos que tomaron. De modo similar, una *relación cuaternaria* entre cuatro conjuntos A, B, C y D está definida como un subconjunto de $((A \times B) \times C) \times D$. En general, una *relación n-aria* entre los conjuntos $A_1, A_2, A_3, \ldots, A_n$ está definida como un subconjunto de $((A_1 \times A_2) \times A_3) \ldots \times A_n$. En otras palabras, una relación n-aria entre los conjuntos A_h $A_2, A_3, \ldots, A_n$ es un conjunto de rc-adas ordenadas en las cuales la primera componente es un elemento de A_x, la segunda componente es un elemento de $A_2, \ldots$, y la w-ésima componente es un elemento de A_n.

4.2 UN MODELO RELACIONAL PARA BASES DE DATOS

Como ejemplo de relaciones entre objetos discretos, presentamos una breve introducción a un modelo relacional para bases de datos. Los modernos sistemas computacionales de gran escala son capaces de manejar grandes cantidades de datos, como los registros de los cargos de las cuentas de los clientes en una tienda departamental, el historial de empleos y datos personales de los empleados en una compañía y los registros de piezas ordenadas y recibidas por una fábrica. Puesto que las grandes cantidades de datos pueden manejarse con eficacia, deben organizarse en una forma que sea adecuada para las operaciones más frecuentes, como la inserción de datos nuevos, eliminación de datos viejos, actualización de los datos existentes y búsqueda para entradas con atributos especiales. Una manera general de considerar la organización de grandes variedades de datos es con un *modelo relacional de datos*. Si $A_1, A_2, \ldots, A_n$ son n conjuntos (no necesariamente distintos), una relación n-aria

SUMINISTROS

PROVEEDOR	PARTE	PROYECTO	CANTIDAD
s_1	p_2	j_5	5
s_1	p_3	j_5	17
s_2	p_3	j_3	9
s_2	p_1	j_5	5
s_4	p_1	j_1	4

Figura 4.4

entre $A_1, A_2, \ldots, A_n$ se conoce como una *tabla*† sobre A_1 $A_2, \ldots, A_n$ en el lenguaje de modelos relaciónales de datos. Los conjuntos $A_1, A_2, \ldots, A_n$ son llamados los *dominios* de la tabla y n es el *grado* de la tabla. Por ejemplo, sean PROVEEDOR = $\{s_1, s_2, s_3, s_4\}$ el conjunto de proveedores de partes, PARTE = $\{p_1,p_2,p_3,p_4,p_5,p_6,p_7\}$ el conjunto de partes, PROYECTO = $\{j_1,j_2,j_3,j_4,j_5\}$ el conjunto de proyectos y CANTIDAD el conjunto de enteros positivos. Podemos tener una tabla llamada SUMINISTROS sobre los conjuntos PROVEEDOR, PARTE, PROYECTO y CANTIDAD, si especificamos los nombres de los proveedores que suministran las partes para los diferentes proyectos y las cantidades que suministran. Así,

$$\text{SUMINISTROS} = \{(s_1, p_2, j_5, 5), (s_1, p_3, j_5, 17), (s_2, p_3, j_3, 9), (s_2, p_1, j_5, 5), (s_4, p_1, j_1, 4)\}$$

es un ejemplo de una tabla. También representaremos una tabla en forma tabular, como se muestra en la figura 4.4.

Otro ejemplo: consideremos una tabla ENSAMBLE sobre PARTE, PARTE y CANTIDAD, donde PARTE = $\{p_1\, p_2, \ldots, p_7\}$ es el conjunto de partes y CANTIDAD es el conjunto de enteros positivos. Una tripleta ordenada (p_i, p_j, c) en la tabla ENSAMBLE significa que la parte/?, es un subcomponente de la parte p_j, además, se necesitan c unidades de la parte p_i para ensamblar cada unidad de la parte p_j. Luego entonces, podríamos tener la tabla mostrada en la figura 4.5.

Un dominio de una tabla es llamado una *llave primaria*‡ si su valor en una n-ada

ENSAMBLE

PARTE	PARTE	CANTIDAD
p_1	p_5	9
p_2	p_5	7
p_3	p_5	2
p_2	p_6	12
p_3	p_6	3
p_4	p_7	1
p_6	p_7	1

Figura 4.5

† El término *relación* también es utilizado. Aquí escogimos utilizar un término más intuitivo.

‡ Debido a que, en general, las n-adas ordenadas se agregan y eliminan de una tabla periódicamente, ser una llave primaria podría ser una propiedad que *varía en el tiempo* para un dominio.

EMPLEADOS

EMPLEADO NUM.	NOMBRE	DEPARTAMENTO	SUELDO POR HORA
20835	Bernstein, E.	2	5.00
11273	Jones, D. J.	1	7.00
10004	Smith, C. W.	1	7.35
21524	Vbgeli, W. J.	2	8.00
17734	Wong, J. W. S.	1	5.00
30219	Yamamoto, S.	3	6.50

Figura 4.6

ordenada identifica únicamente la n-ada ordenada en la tabla. En la tabla EMPLEADOS en la figura 4.6, EMPLEADO NÚM. es una llave primaria, como también lo es NOMBRE. Por otro lado, DEPARTAMENTO no es una llave primaria y tampoco lo es SUELDO POR HORA. Si una tabla no tiene un dominio que pueda servir como llave primaria, podríamos pensar en usar una combinación de dominios para identificar las n-ada ordenadas de la tabla. Definimos una *llave primaria compuesta* como el producto cartesiano de dos o más dominios tal que su valor[†] en una n-ada ordenada identifique únicamente la n-ada ordenada en la tabla. Por ejemplo, en la tabla SUMINISTROS en la figura 4.4, PROVEEDOR X PARTE es una llave primaria compuesta.

Tenemos una colección de tablas y quisiéramos manipularlas de varias maneras. Como un ejemplo, describimos dos importantes operaciones sobre tablas: *proyección* y *conjunción.* Con frecuencia queremos extraer subtablas a partir de una tabla. La operación de *proyección* nos permite efectuar dicha extracción. Si R es una tabla de grado n, una *proyección* de R es una relación m-aria, $m \leqq n$, obtenida a partir de R al eliminar $n - m$ de las componentes en cada n-ada ordenada en R. Usamos la notación $\pi_{i_1 i_2 \ldots i_m}(R)$, $1 \leqq i_1 < i_2 < \cdots < i_m \leqq n$, para denotar una proyección de R, esto es, una tabla de grado m obtenida a partir de R tal que para cada n-ada ordenada en R hay una correspondiente m-ada ordenada en $\pi_{i_1 i_2 \ldots i_m}(R)$ donde la k-ésima componente de la m-ada ordenada es la i_k-ésima componente de la n-ada ordenada. Por ejemplo, para la tabla SUMINISTROS en la figura 4.4, la proyección $\pi_{1,3}$ (SUMINISTROS) se muestra en la figura 4.7.

Observemos que, como se ilustró en este ejemplo, podría haber menos m-ada ordenadas en una proyección de una tabla que las «-ada ordenadas que hay en la tabla, debido a que varias n-ada ordenadas distintas en la tabla podrían dar origen a la misma m-ada ordenada en la proyección.

$\pi_{1,3}$(SUMINISTROS)

PROVEEDOR	PROYECTO
s_1	j_5
s_2	j_3
s_2	j_5
s_2	j_1

Figura 4.7

[†] Por los valores de una llave primaria compuesta, entendemos las n-adas ordenadas en el producto cartesiano.

SUMINISTROS

PROVEEDOR	PARTE	PROYECTO
s_1	p_1	j_1
s_2	p_1	j_1
s_2	p_2	j_2

COLOR

PARTE	PROYECTO	COLOR
p_1	j_1	c_1
p_2	j_2	c_2
p_2	j_2	c_3

a)

τ_2 **(SUMINISTROS * COLOR)**

PROVEEDOR	PARTE	PROYECTO	COLOR
s_1	p_1	j_1	c_1
s_2	p_1	j_1	c_1
s_2	p_2	j_2	c_2
s_2	p_2	j_2	c_3

Figura 4.8 *b*)

Las tablas también pueden combinarse para dar origen a tablas más grandes. La operación de *conjunción* combina dos tablas en una sola. Sean R una tabla de grado n y S una tabla de grado m. Para una p menor que n y m, podemos construir una *conjunción* de R y S, es decir, una tabla denotada por $\tau_p(R * S)$ tal que

$$\begin{aligned}\tau_p(R * S) = \{(a_1, a_2, \ldots, a_{n-p}, b_1, b_2, \ldots, b_p, c_1, c_2, \ldots, c_{m-p})\\ (a_1, a_2, \ldots, a_{n-p}, b_1, b_2, \ldots, b_p) \in R,\\ (b_1, b_2, \ldots, b_p, c_1, c_2, \ldots, c_{m-p}) \in S\}\end{aligned}$$

Para ejemplificar con las tablas SUMINISTROS y COLOR de la figura 4.8a, tenemos la conjunción τ_2 (SUMINISTROS * COLOR) mostrada en la figura 4.86.

Se puede decir mucho más acerca del modelo relacional de datos, en especial en lo referente a otras operaciones sobre tablas, su introducción en sistemas computacionales, etcétera; muchos de estos detalles pueden encontrarse en Codd [1] y Date [3].

4.3 PROPIEDADES DE LAS RELACIONES BINARIAS

Una relación binaria de un conjunto A sobre A se dice que es una *relación binaria sobre A*. Por ejemplo, si A es un conjunto de enteros positivos, podemos definir una relación binaria R sobre A tal que (a, b) está en R si y sólo si $a - b \geqq 10$. Así (12, 1) está en R, pero (12, 3) no lo está; como tampoco está (1,12). Otro ejemplo, sea 5 = {*CS*121, *CS*221, *CS*257, *CS*264, *CS*273,*CS*281}. Una relación binaria sobre B={(*CS*121,*CS*221),(*CS*121,*CS*257),(*CS*257, *CS*281)} podría describir la estructura de prerrequisitos de estos cursos de manera que un par ordenado en la relación binaria significa que el primer curso en el par es un prerrequisito del segundo curso en el par. En lo que resta de este capítulo, estudiaremos las relaciones

	a	b	c	d
a	√			√
b		√		√
c			√	
d				√

a)

	a	b	c	d
a	√			√
b		√		√
c				
d				√

b)

Figura 4.9

binarias sobre un conjunto, debido a que éste es el caso en que las encontraremos con mayor frecuencia.

Sea *R* una relación binaria sobre *A*. Diremos que *R* es una *relación reflexiva* si *(a, a)* está en *R* para todo *a* en *A*. En otras palabras, en una relación reflexiva todo elemento de *A* está relacionado con él mismo. Por ejemplo, si *A* es un conjunto de cursos y *R* es una relación binaria sobre *A* de manera que para dos cursos *a* y *b* en *A*, *(a, b)* está en *R* si y sólo si sus evaluaciones finales están programadas en el mismo periodo de tiempo. Es evidente que para un curso cualquiera *a*, *(a, a)* está en *R*. Así, *R* es una relación reflexiva. Como otro ejemplo, sea *A* un conjunto de enteros positivos, y definimos una relación binaria *R* sobre *A* tal que *(a, b)* está en *R* si y sólo si *a* divide a *b*. Debido a que un entero siempre se divide él mismo, *R* es una relación reflexiva. Por otro lado, si definimos una relación binaria *T* sobre un conjunto de enteros *A* de manera que *(a, b)* está en *T* si y sólo si $a > b$. Es evidente que *T* no es una relación reflexiva. En otro ejemplo sea *A* un conjunto de estudiantes y *R* una relación binaria sobre *A* tal que *(a, b)* está en *R* si *a* propone a *b* como candidato para presidente de la generación. *R* es una relación reflexiva si *cada uno* se propone a sí mismo. Por otro lado, *R* no es una relación reflexiva si uno o más de los estudiantes no lo hiciera. Cuando una relación binaria sobre un conjunto es representada en forma tabular, es muy sencillo determinar si la relación binaria es una relación reflexiva. Para ser específicos, una relación binaria sobre un conjunto es reflexiva si y sólo si todas las celdas sobre la diagonal principal de la tabla contienen marcas. Por ejemplo, la relación binaria en la figura 4.9*a* es reflexiva, en tanto que la de la figura 4.9*b* no lo es.

Sea *R* una relación binaria sobre *A*. Diremos que *R* es una *relación simétrica* si *{a, b)* en *R* implica que *(b, a)* también está en *R*. Por ejemplo, sean *A* un conjunto de estudiantes y *R* una relación binaria sobre *A* tal que *(a, b)* está en *R* si y sólo si *a* está en una clase en la que *b* se encuentra. Si *a* está en una clase en la que *b* se encuentra, entonces, es claro que *b* también está en la clase en la que *a* se encuentra. Así, la relación *R* es una relación simétrica. Sean *A* un conjunto de enteros positivos y *T* una relación binaria sobre *A* tal que *(a, b)* está en *T* si y sólo si $a \geqq b$. Por ejemplo, como (10,9) está en *T* pero (9, 10) no está en *T*, *T* no es una relación simétrica. Como otro ejemplo, sea $A = \{a, b, c\}$, entonces $U = \{\}$, $V = \{(a, a), (b, b)\}$, $W = \{\{a, b), (b, a)\}$ y $X = A \times A$ son cuatro relaciones binarias sobre *A*. Observemos que las cuatro son relaciones simétricas. Cuando una relación binaria sobre un conjunto se representa en forma tabular, podemos determinar si ésta es simétrica al observar si las marcas están en celdas que son simétricas con respecto de la diagonal principal. Por ejemplo, la re-

	a	b	c	d
a		√		√
b	√		√	
c		√		√
d	√		√	√

a)

	a	b	c	d
a		√		√
b	√			
c				
d				√

b)

Figura 4.10

lación binaria en la figura 4.10a es simétrica, en tanto que la de la figura 4.10*b* no es simétrica.

Sea *R* es una relación binaria sobre *A.* Diremos que *R* es una *relación antisimétrica* si *(a, b)* en *R* implica que *(b, a)* no está en *R* a menos que *a* = *b.* En otras palabras, si tanto *(a, b)* como *(b, a)* están en *R*, entonces se debe dar el caso de que *a = b.* Por ejemplo, si *A* es el conjunto de exámenes que se realizan a un paciente en el hospital y *R* es una relación binaria sobre *A* tal que si *(a, b)* está en *R,* entonces el examen *a* debe realizarse antes del examen *b.* Es obvio que si el examen *a* debe realizarse antes que el examen *b,* entonces el examen *b* no debe realizarse antes del examen *a* para dos exámenes distintos *ay b*. Luego, *R* es una relación antisimétrica. Otro ejemplo, sea ,4 un conjunto de enteros positivos y *R* una relación binaria sobre *A* tal que *(a, b)* está en *R* si y sólo si $a \geqq b$. Observamos que *R* es una relación antisimétrica. Sea *A* = *{a, b, c}.* Si *S= {(a, a), {b, b)} yN= {(a, b), (a, c), (c, a)}* son relaciones binarias sobre *A,* observamos que *S* es tanto simétrica como antisimétrica, cuando *N* no es ni simétrica ni antisimétrica.

Sea *R* una relación binaria sobre *A.* Diremos que *R* es una *relación transitiva* si (a, c) está en *R* siempre que, tanto *(a, b)* como (6, c) estén en *R*. Por ejemplo, si *A* = *{a, b, c}* y *X*= *{(a, a), {a, b), (a, c), (b,* c)}, observamos quedes una relación transitiva. También notamos que *Y- {{a, b)}* es una relación transitiva, mientras que Z = *{(a, b),* (¿>, c)} no lo es.[†] Otro ejemplo es, si *A* es un conjunto de personas y *R* es una relación binaria sobre *A* tal que (a, *b)* está en *R* si y sólo si *a* es un antepasado de *b.* Es claro que *R* es una relación transitiva. Por otro lado, si tenemos una relación binaria *T* sobre *A* tal que *{a, b)* está en *T* si y sólo si *a* es el padre de *b,* entonces *Tno* es una relación transitiva.

Sea *R* una relación binaria sobre *A.* La *extensión transitiva de R,* denotada por R_x, es una relación binaria sobre *A* tal que R_l contiene *aRy* además si *(a, b) y (b, c)* están en *R,* entonces *(a, c)* está en R_v Por ejemplo, *sea A* = {a, *b, c, d) y R* la relación binaria mostrada en la figura 4.1 la. La extensión transitiva de *R*, R_l, se muestra en la figura 4.11 *b,* donde los pares ordenados en R_1 pero no en *R,* están marcados con marcas en negrita. Observemos que si *R* es una relación transitiva, *R* es igual a R_1. Si R_2 denota la extensión transitiva de

[†] En este momento, es probable que el lector se dará cuenta del problema de verificar cuándo una determinada relación binaria es transitiva o no. Esto puede realizarse mediante una búsqueda exhaustiva, no obstante que la búsqueda de procedimientos eficientes para la verificación de la transitividad es aún un tópico actual de investigación.

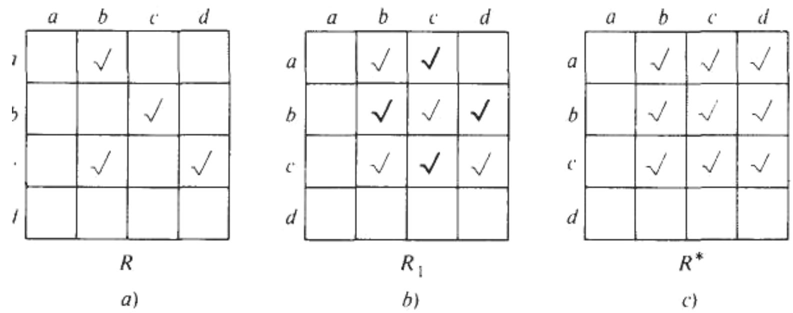


Figura 4.11

R_1 y, en general, R_{i+1} denota la extensión transitiva de R_i definimos la *cerradura transitiva* de R, denotada por R^*, como el conjunto unión de R, R_1 R_2,. .. Por ejemplo, la cerradura transitiva de la relación binaria R en la figura 4.11*a* se muestra en la figura 4.11*c*. Otro ejemplo: consideremos que A es un conjunto de ciudades *y Res* una relación binaria sobre A de manera que el par ordenado (a, b) está en R si existe un enlace de comunicación desde la ciudad a hasta la ciudad b para la transmisión de mensajes. Así, la extensión transitiva de R, R_1 describe cómo los mensajes pueden ser transmitidos desde una ciudad a otra, ya sea por un enlace de comunicación directa o a través de una ciudad intermedia. De modo similar, la extensión transitiva de R_1, R_2, describe cómo un mensaje puede ser transmitido de una ciudad a otra, mediante un enlace de comunicación directa o a través de cuanto más tres ciudades intermedias.† Por último, la cerradura transitiva de R, R^*, describe cómo pueden ser transmitidos mensajes desde una ciudad hasta otra, por medio de un enlace de comunicación directo o a través del número que se quiera de ciudades intermedias.

Sean A un conjunto de hombres y R una relación binaria sobre A, de manera que el par ordenado (a, *b)* está en R si a es el padre de b. En general, R no es una relación transitiva puesto que si a es el padre de b y b es el padre de c, entonces, en definitiva a no es el padre de c. Por otro lado, la cerradura transitiva de R, R^*, es una relación transitiva que describe la relación antepasado-descendiente entre las personas en A. Observemos que para cualquier relación binaria R, R^* siempre es una relación transitiva.

4.4 RELACIONES DE EQUIVALENCIA Y PARTICIONES

Una relación binaria puede tener una o más de las siguientes propiedades: reflexividad, simetría, antisimetría y transitividad. Por ejemplo, la relación binaria en la figura 4.12*a* es una relación reflexiva y transitiva, y la relación binaria en la figura 4.126 es una relación reflexiva y simétrica. En esta sección y la siguiente estudiaremos dos clases importantes de relaciones binarias, a saber, las relaciones de equivalencia y las relaciones de orden parcial.

---

†Véase el problema 4.23.

	a	b	c	d
a	√	√		√
b		√		√
c			√	
d		√		√

a)

	a	b	c	d
a	√	√		
b	√	√	√	
c		√	√	
d				√

b)

Figura 4.12

Diremos que una relación binaria sobre un conjunto es una *relación de equivalencia* si ésta es reflexiva, simétrica y transitiva. Por ejemplo, la relación binaria sobre el conjunto *{a, b, c, d, e, f}* mostrada en la figura 4.13 es una relación de equivalencia. Sean *A* un conjunto de estudiantes y *R* una relación binaria de *A* tal que *(a, b)* está en i? si y sólo si *a* vive en el mismo dormitorio que *b.* Puesto que cualquier persona vive en el mismo dormitorio que ella misma, *R* es una relación reflexiva. Si *a* vive en el mismo dormitorio que *b,* entonces *b* vive en el mismo dormitorio que *a.* Así, *R* es una relación simétrica. Observemos que si *a* vive en el mismo dormitorio que *b* y *b* vive en el mismo dormitorio que c, entonces *a* vive en el mismo dormitorio que *c.* Por consiguiente, *R* es una relación transitiva. En consecuencia, *R* es una relación de equivalencia. Ahora consideremos que *A* es un conjunto de cadenas de ceros y unos cuyas longitudes son al menos de tres dígitos. Sea *R* una relación binaria sobre *A* tal que para las cadenas *a* y *b {a, b)* está en *R* si y sólo si los tres últimos dígitos en *a* son los mismos que los tres últimos dígitos en *b.* Dejamos al lector la verificación de que *R* es una relación de equivalencia. Intuitivamente, en una relación de equivalencia dos objetos están relacionados si comparten algunas propiedades en común o si satisfacen algunos requerimientos en común, y entonces son "equivalentes" con respecto a estas propiedades o requerimientos.

Ahora definimos el concepto *de partición* de un conjunto. Una *partición* de un conjunto *A* es un conjunto de subconjuntos no vacíos de *A* denotada por $\{A_1\ A_2,. . ., A_k\}$ tal que la unión de los A_i es igual a *A*, y la intersección de A_i con A_j es vacía para cualesquiera A_i y *Aj* distintos. En otras palabras, una partición de un conjunto es una división de los elementos del conjunto en subconjuntos disjuntos. Estos subconjuntos también se llaman *bloques* de

	a	b	c	d	e	f
a	√	√				
b	√	√				
c			√			
d				√	√	√
e				√	√	√
f				√	√	√

Figura 4.13

la partición. Por ejemplo, si $A = \{a, b,c, d,e,f,g\}$, entonces $\{\{a\}, \{b,c,d\}, \{e,f\} \{g\}\}$ es una partición de A. En otro ejemplo, observemos que una baraja de cartas está particionada en cuatro figuras, así como también está particionada en 13 niveles. Introducimos la notación $\{\overline{a}, \overline{bcd}, \overline{ef}, \overline{g}\}$ para una partición, en la que colocamos una barra sobre los elementos que están en el mismo bloque. Ahora veremos una conexión entre una relación de equivalencia sobre un conjunto A y una partición del conjunto A. A partir de esta relación de equivalencia sobre A, podemos definir una partición de A, de modo que dos elementos cualesquiera de un bloque están relacionados y cualesquiera dos elementos de bloques diferentes no están relacionados.† Se dice que esta partición es una partición inducida por la relación de equivalencia, y los bloques de la partición se conocen como *clases de equivalencia*. Recíprocamente, a partir de una partición de un conjunto A podemos definir una relación de equivalencia sobre A de manera que cualesquiera dos elementos en el mismo bloque de la partición están relacionados, y cualesquiera dos elementos en diferentes bloques no están relacionados.‡ Por ejemplo, si A es un conjunto de personas y R es una relación binaria sobre A tal que (a, b) está en R si y sólo si *a y b* tienen el mismo apellido familiar, nos damos cuenta que R es una relación de equivalencia, la cual induce una partición de A, donde las clases de equivalencia son las familias.§ En otro ejemplo: consideremos que A es el conjunto de todos los números naturales. Sea n un entero fijo. Si R es una relación binaria sobre A de manera que (a, b) está en R si los residuos de a dividido por n y b dividido por n son iguales. R es una relación de equivalencia que divide A en n clases de equivalencia. Estas clases de equivalencia son aquellos números divisibles por n, aquellos números que dejan un residuo de 1 al dividirlos por n, aquellos números cuyo residuo es 2 cuando se dividen por $n, \ldots,$ y aquellos números que dejan un residuo de $n - 1$ cuando se dividen por n. Diremos que dos números a y b que se encuentran en la misma clase de equivalencia son *igual módulo n*, y la notación $a \equiv b \pmod{n}$ es usada frecuentemente.

Sean π_1 y π_2 dos particiones de un conjunto A, y sean R_1 y R_2 sus relaciones de equivalencia respectivas. Diremos que π_1 es un *refinamiento* de π_2, denotado por $\pi_1 \leq \pi_2$, si $R_1 \subseteq R_2$. En otras palabras, si π_1 es un refinamiento de π_2, entonces cualesquiera dos elementos que están en el mismo bloque de π_1, también deben estar en el mismo bloque de π_2. Definimos el *producto* de π_1 y π_2, denotado $\pi_1 \cdot \pi_2$, como la partición correspondiente a la relación de equivalencia $R_1 \cap R_2$ (el lector demostrará que la intersección de dos relaciones de equivalencia siempre es una relación de equivalencia). En otras palabras, el producto de π_1 y π_2 es una partición de A tal que dos elementos *a y b* están en el mismo bloque de $\pi_1 \cdot \pi_2$ si a y b están en el mismo bloque de n_x *y* también en el mismo bloque de π_2. Así, $\pi_1 \cdot \pi_2$ es un refinamiento de π_1 y asimismo un refinamiento de π_2. Definimos la *suma* de π_1 y π_2, denotada $\pi_1 + \pi_2$, como la partición correspondiente a la relación de equivalencia $(R_1 \cup R_2)^*$ (el lector demostrará que la unión de dos relaciones de equivalencia siempre es una relación reflexiva y simétrica). En otras palabras, la suma de π_1 y π_2 es una partición de A tal que dos elementos *a y b* están en el mismo bloque de $\pi_1 + \pi_2$ si existen elementos $c_1, c_2, c_3, \ldots, c_k$ tales que a y c_1 están en el mismo bloque de π_1 o π_2, c_1 y c_2 están en el mismo bloque de π_1 o $\pi_2, \ldots, c_2$ y c_3 están en el mismo bloque de π_1 o $\pi_2, \ldots,$ y c y b están en el

† Véase el problema 4.24, inciso a).
‡ Véase el problema 4.24, inciso *b)*.
§ Suponemos que no hay dos familias con el mismo apellido.

mismo bloque de π_1 o π_2. Esto es, dos elementos *a y b* están en el mismo bloque de $\pi_1 + \pi_2$ si se encuentran *conectados en cadena,* con lo cual queremos señalar que existe una sucesión de elementos $a, c_1, c_2, \ldots, c_k, b$ tales que cada par de elementos sucesivos en la sucesión están en el mismo bloque de π_1 o π_2. Así, tanto π_1 como π_2 son refinamientos de $\pi_1 + \pi_2$.

Por ejemplo, sea $A = \{a, b, c, d, e, f, g, h, i, j, k\}$. Y sean

$$\pi_1 = \{\overline{abcd}, \overline{efg}, \overline{hi}, \overline{jk}\} \qquad \pi_2 = \{\overline{abch}, \overline{di}, \overline{efjk}, \overline{g}\}$$

dos particiones de *A*. Tenemos entonces que

$$\pi_1 \cdot \pi_2 = \{\overline{abc}, \overline{d}, \overline{ef}, \overline{g}, \overline{h}, \overline{i}, \overline{jk}\}$$

y

$$\pi_1 + \pi_2 = \{\overline{abcdhi}, \overline{efgjk}\}$$

Una interpretación física interesante puede darse para el producto y la suma de particiones. Para el conjunto *A y* las particiones π_1 y π_2 dados arriba, suponemos que *A* es un conjunto de personas, π_1 es una partición de ellas en grupos de edad, y π_2 es una partición de ellas en grupos de estatura. Supongamos que deseamos identificar a cierta persona en *A.* Si se nos informa en qué grupo de edad se encuentra, entonces podemos identificarla a partir de los bloques en π_1. En otras palabras, debemos ser capaces de identificarla como una de *a, b, c, d,* o como una de *e,f,g,* o como una de *h, i,* o como una de *j, k,* según sea el grupo de edad en el que se encuentra. De manera similar, si nos indican en qué grupo de estatura está la persona, podemos identificarla a partir de los bloques en π_2. Si nos indican el grupo de edad y el grupo de estatura en los que se encuentra, entonces podemos identificarla a partir de los bloques en $\pi_1 \cdot \pi_2$. En otras palabras, debemos identificarla como una de *a, b, c,* como *d,* como una de *e,f,* como *g,* como *h,* como *i,* o como una de *j, k.* Si se nos indica el grupo de edad o el de estatura en el que se encuentra, pero no ambos, podemos identificarla a partir de los bloques en $\pi_1 + \pi_2$. En otras palabras, ya sea que se nos proporcione información sobre el grupo de edad o información sobre el grupo de estatura en el que se encuentra la persona a identificar, no habrá ambigüedad alguna sobre la identidad de la persona más allá de los bloques de $\pi_1 + \pi_2$. Es decir, estamos seguros que podemos distinguir a alguien en el grupo *a, b, c, d, h, i* de otra persona del grupo *e, f, g, j, k,* sin importar si se nos da información sobre el grupo de edad o del grupo de estatura. En efecto, una partición sobre un conjunto puede ser vista como la posesión de alguna *información* sobre la identificación de uno de los objetos en el conjunto (dos objetos en bloques diferentes siempre podrán distinguirse), o como la posesión de alguna *ambigüedad* sobre la identificación de uno de los objetos en el conjunto (dos objetos en el mismo bloque nunca podrán distinguirse). De esto se sigue que el producto de dos particiones $\pi_1 \cdot \pi_2$, representa la información total que tenemos sobre la identificación de uno de los objetos cuando tenemos la información tanto para π_1 como para π_2, y la suma de las dos particiones $\pi_1 + \pi_2$ representa la *mayor*† ambigüedad que podríamos tener cuando estamos seguros de que dispondremos solamente de información ya sea para π_1 o bien para π_2.

---

† Observe el uso de la palabra *mayor.* En muchos casos, la ambigüedad puede ser menor que la que hay en $\pi_1 + \pi_2$.

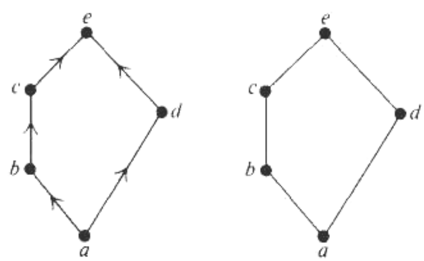
4.5 RELACIONES DE ORDEN PARCIAL Y LATTICES†

Diremos que una relación binaria es una *relación de orden parcial* si ésta es reflexiva, antisimétrica y transitiva. Por ejemplo, la relación binaria sobre el conjunto *{a, b, c, d, e}* mostrada en la figura 4.14*a* es una relación de orden parcial. En otro ejemplo, sea *A* un conjunto de enteros positivos, y sea *R* la relación binaria sobre *A* tal que *(a, b)* está en *R* si *a* divide a *b.* Debido a que cualquier entero se divide a sí mismo, *R* es una relación reflexiva. Si *a* divide a *b* significa que *b* no divide a a, a menos que *a=b, R* es una relación antisimétrica. Puesto que si *a* divide a *b y b* divide a c, entonces *a* divide a c, y *R* es una relación transitiva. En consecuencia, *R* es una relación de orden parcial. Consideremos también el ejemplo de un conjunto de libros, *A,* cada uno de los cuales posee un cierto número de atributos. Sea *R* una relación binaria sobre *A* tal que *(a, b)* está en *R* si y sólo si cada atributo del libro *a* es también un atributo del libro *b. R* es una relación de orden parcial. Otro ejemplo: sea *A* un conjunto de productos alimenticios de diferentes precios, y sea *R* una relación binaria sobre *A* tal que *(a, b)* está en *R* si *a* no es inferior a *b* en términos tanto de valor nutricional como de precio. Otra vez, *R* es una relación de orden parcial. Intuitivamente, en una relación de orden parcial dos objetos están relacionados si uno de ellos es menor (mayor) que, o inferior (superior) al otro objeto de acuerdo con algunas propiedades o criterios. En efecto, la palabra *orden* implica que los objetos en el conjunto están ordenados de acuerdo con estas propiedades o criterios. No obstante, también es posible que dados dos objetos en el conjunto, éstos no tengan la relación de orden parcial. En tal caso, no podemos comparar estos dos objetos e identificar el menor o el inferior. Ésta es la razón de usar el término *orden parcial.*

En la sección 4.1 se señaló que una relación binaria de un conjunto *A* a un conjunto *B* puede ser representada gráficamente como se ilustra en la figura 4.1*b*. Para una relación binaria *R* sobre un conjunto *A,* podemos tener una representación gráfica un poco más simple. En lugar de tener dos columnas de puntos como en la figura 4.1*b*, representamos los elementos de *A* por puntos y usamos flechas para representar los pares ordenados en *R.* Por ejemplo, la relación binaria sobre el conjunto *{a, b, c, d}* en la figura 4.15*a* está representada en una gráfica en la figura 4.15*b*. Cuando la relación binaria es una relación de orden parcial, la representación gráfica puede simplificarse aún más. Puesto que la relación es reflexiva, podemos omitir flechas de puntos que regresen a ellos mismos. Ya que la relación es tran-

	a	*b*	*c*	*d*	*e*
a	✓	✓	✓	✓	✓
b		✓	✓		✓
c			✓		✓
d				✓	✓
e					✓

a)



b) *c*)

Figura 4.14

† *N. del E.* Se prefirió respetar la grafía original de la palabra *lattice*, aun cuando en contexto matemático, en español equivale a celosía, retícula o reticulado.

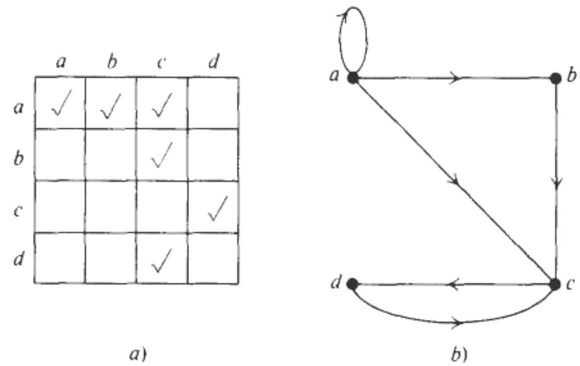


Figura 4.15

sitiva, podemos omitir las flechas entre puntos que están conectados por sucesión de flechas. Por ejemplo, tal representación simplificada para la relación de orden parcial en la figura 4.14*a* se muestra en la figura 4.14*b*. En muchos casos, cuando la representación gráfica está muy orientada, de manera que todas las puntas de flecha apuntan en la misma dirección (hacia arriba, hacia abajo, de izquierda a derecha, o de derecha a izquierda), aún podemos omitir las puntas de flecha como se muestra en el ejemplo de la figura 4.14*c*. La representación gráfica de una relación de orden parcial en la cual todas las puntas de flecha apuntan hacia arriba se conoce también como *diagrama de Hasse* de la relación.

Un conjunto *A*, junto con la relación de orden parcial *R* sobre *A*, es llamado un *conjunto parcialmente ordenado* y se denota por *(A, R)*. En la literatura, un conjunto parcialmente ordenado también se conoce como *poseí* (del inglés: *p*artially ordered *set)*. También existe otra notación para especificar una relación de orden parcial: para cada par ordenado *(a, b)* en *R*, escribimos $a \le b$ en vez de $(a, b) \in R$, donde ≤ es un símbolo genérico que corresponde al conjunto de pares ordenados *R*, y por lo común se lee "menor o igual que" (con frecuencia decimos *a* es menor o igual que *b* para entender que $a \le b$, y decimos *a* es menor que *b*, para entender que $a \le b$ y $a \ne b$. También decimos *b* es mayor o igual que *a* y escribimos $b \ge a$ para entender que $a \le b$). En realidad, un conjunto parcialmente ordenado se denota como $(A, \le)$.

Sea $(A, \le)$ un conjunto parcialmente ordenado. Un subconjunto de *A* se llama una *cadena* si cualesquiera dos elementos en el subconjunto están relacionados. Observemos que, a causa de la antisimetría y la transitividad, en cualquier cadena con un número finito de elementos $\{a_1, a_2, \ldots, a_k\}$ existe un elemento a_{i_1}, que es menor que cualquier otro elemento de la cadena, existe un elemento a_{i_2} que es menor que cualquier otro elemento excepto a_{i_1}, existe un elemento a_{i_3} que es menor que cualquier otro elemento excepto a_{i_1}, y a_{i_2} y así sucesivamente. Usaremos la notación $a_{i_1} \le a_{i_2} \le a_{i_3} \le \cdots \le a_{i_k}$ como una abreviatura de la lista de pares ordenados $a_{i_1} \le a_{i_2}, a_{i_1} \le a_{i_3}, \ldots, a_{i_1} \le a_{i_k}, a_{i_2} \le a_{i_3}, a_{i_2} \le a_{i_4}, \ldots, a_{i_2} \le a_{i_k}, \ldots$ Con frecuencia nos referiremos al número de elementos en la cadena como la *longitud* de la cadena. Un subconjunto de *A* es llamado una *anticadena* si no hay dos elementos distintos en el subconjunto que estén relacionados. Por ejemplo, para el conjunto parcialmente ordenado de la figura 4.14a, {a, *b, c, d)*, {*a, b, c}, {a, d, e) y {a}* son cadenas, y *{b, d), {c, d)* y *{a}* son anticadenas. Consideremos un conjunto parcialmente ordenado $(A, \le)$, donde *A* es el conjunto de todos los empleados de una compañía, y para *a* y *b* en *A*, $a \le b$ si y sólo si *b* es a o es un superior de a. En este caso, una cadena es un subconjunto de empleados en el cual,

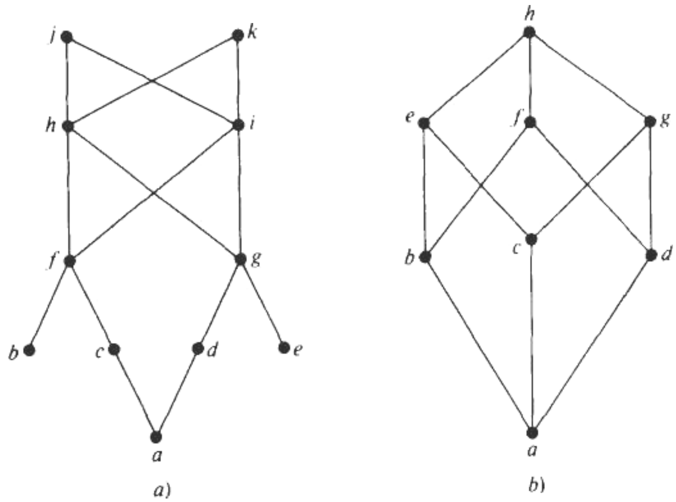


Figura 4.16

en efecto, existe una cadena de mando. Por otro lado, una anticadena es un subconjunto de empleados en la cual ninguno tiene mando sobre otro. Un conjunto parcialmente ordenado $(A, \leq)$ es llamado un *conjunto totalmente ordenado* si *A* es una cadena. En este caso, la relación binaria $\leq$ es llamada una *relación de orden total.*

Sea $(A, \leq)$ un conjunto parcialmente ordenado. Un elemento *a* en *A* es llamado un *elemento maximal* si para ningún *b* en *A*, $a \neq b$, $a \leq b$. Un elemento *a* en *A* es llamado un *elemento minimal* si para ningún *b* en A, $a \neq b$, $b \leq a$. Por ejemplo, en el conjunto parcialmente ordenado de la figura 4.16*a*, *j y k* son elementos maximales, y *a, b, e* son elementos minimales. Diremos que un elemento *a cubre* a otro elemento *b* si $b \leq a$ y no hay otro elemento *c*, tal que $b \leq c \leq a$. En el conjunto parcialmente ordenado de la figura 4.16*a*, *f* cubre a *b, f* también cubre a *c*, pero *f* no cubre a *a*.

Sean *a* y *b* dos elementos en un conjunto parcialmente ordenado $(A, \leq)$. Diremos que un elemento *c* es una *cota superior* de a y b si $a \leq c$ y $b \leq c$. Por ejemplo, en el conjunto parcialmente ordenado de la figura 4.16*a*, *h* es una cota superior de *f* y *g*; también lo son *i*, *j* y *k*. Diremos que un elemento c es una *cota superior mínima* de *a* y *b* si *c* es una cota superior de *a* y *b, y* no existe otra cota superior *d* de *a* y *b* tal que $d \leq c$. Por ejemplo, en el conjunto parcialmente ordenado de la figura 4.16*a*, *h* es una cota superior mínima de *f* y *g*, como también lo es *i*. De modo similar, diremos que un elemento *c* es una *cota inferior* de a y b si $c \leq a$ y $c \leq b$, y un elemento c es una *cota inferior máxima* de *a* y *b* si *c* es una cota inferior de *a y b, y* no existe otra cota inferior *d* de *a y b* tal que $c \leq d$. En el conjunto parcialmente ordenado de la figura 4.16*a*, los elementos a, *b, c, d, e, f* y *g* son todos cotas inferiores de *h* e *i*, en tanto que *f* y *g* son cotas inferiores máximas de *h* e *i*. En otro ejemplo, sea *A* el conjunto de todos los enteros positivos y *R* una relación binaria sobre *A* tal que (a, *b)* está en *R* si *a* divide a *b.* Podemos comprobar con facilidad que *R* es una relación de orden parcial. Para dos enteros *a* y *b*, un múltiplo común de *a* y *b* es una cota superior de *a* y *b*, y el mínimo común múltiplo de *a* y *b* es una cota superior mínima (y la única) de *a y b.* De manera similar, un divisor de *a* y *b* es una cota inferior de *a* y *b*, y el máximo común divisor de *a* y *b* es una cota inferior máxima (y la única) de *a* y *b.*

Se dice que un conjunto parcialmente ordenado es una *lattice* si cualesquiera dos elementos en el conjunto tienen una cota superior mínima y única, y una cota inferior máxima

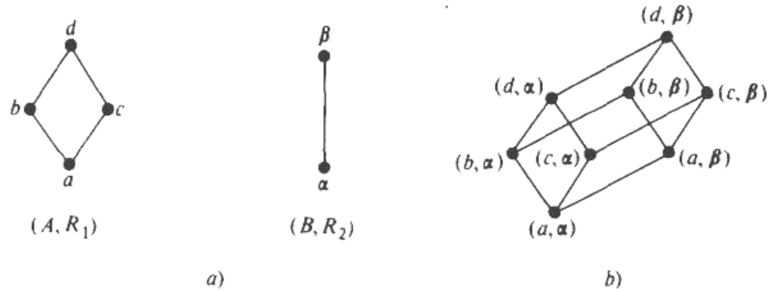


Figura 4.17

única. El conjunto parcialmente ordenado de la figura 4.16*a* no es una lattice, en tanto que sí lo es el de la figura 4.16*b*. Estudiaremos varias propiedades de lattices en el capítulo 12. Sean (A, R_1) y (B, R_2) dos conjuntos parcialmente ordenados. Definimos una relación binaria R_3 sobre el conjunto $A \times B$ tal que para a_1 y a_2 en A, y b_1 y b_2 en B $((a_1, b_1)), (a_2, b_2))$ está en R_3 si y sólo si (a_1, a_2) está en R_1 y $(b_1\ b_2)$ está en R_2. Invitamos al lector a demostrar que R_3 es una relación de orden parcial. Por tanto, $(A \times B, R_3)$ es un conjunto parcialmente ordenado que por lo general se conoce como el *producto cartesiano* de los dos conjuntos parcialmente ordenados (A, R_1) y (B, R_2). La figura 4.17*b* muestra el producto cartesiano de los dos conjuntos parcialmente ordenados en la figura 4.17*a*. Sean *A* el conjunto de todos los divisores positivos de un entero *n*, y *R* una relación binaria sobre *A* tal que (a, *b)* está en *R* si *a* divide a *b*. Si expresamos a *n* como un producto de potencias de números primos $p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_t^{\alpha_t}$, notamos que un divisor positivo de *n* puede ser expresado como $p_1^{\beta_1} p_2^{\beta_2} \cdots p_t^{\beta_t}$, donde $0 \leqq \beta_i \leqq \alpha_i$, para $i = 1, 2, \ldots, t$. En efecto, todos los enteros en *A* pueden ser representados por *t*-adas ordenadas de la forma $(\beta_1, \beta_2, \ldots, \beta_t)$, donde $0 \leqq \beta_i \leqq \alpha_i$, $i = 1, 2, \ldots, t$. Además, sean las í-adas ordenadas correspondientes a los enteros a y b $(\beta_1, \beta_2, \ldots, \beta_t)$ y $(\beta_1', \beta_2', \ldots, \beta_t')$. Que *a* divide a *b* significa que $\beta_i \leq \beta_i'$, $i = 1, 2, \ldots, t$. Dejamos al lector verificar que el conjunto parcialmente ordenado (A, R) puede expresarse como el producto cartesiano $(((A_x \times A_2) \times A_3) \times \ldots) \times A_t$ donde, para $i = 1, 2, \ldots, t$ $(A_i, \leq)$ es un conjunto totalmente ordenado tal que $A_i = \{0, 1, 2, \ldots, \alpha_i\}$ y $0 \leq 1 \leq 2 \leq \cdots \leq \alpha_i$.

4.6 CADENAS YANTICADENAS

Como ilustración de los conceptos de cadenas y anticadenas en conjuntos parcialmente ordenados, consideremos un ejemplo sencillo:

Ejemplo 4.1

Sea $A = \{a_l, a_2, \ldots, a_r\}$ el conjunto de todos los cursos requeridos para la graduación. Sea *R* una relación binaria *reflexiva* sobre *A* tal que para $a_i \neq a_j$, (a_i, a_j) está en i? si y sólo si el curso a_i es un prerrequisito del curso a_j. *R* es antisimétrica y transitiva. En consecuencia, *R* es una relación de orden parcial.† Supongamos que la longitud de la cadena más

† En la literatura una relación antisimétrica y transitiva es referida como una *relación de precedencia*. Las relaciones de precedencia comparten muchas de las propiedades de las relaciones de orden parcial. No obstante, cuando una situación física conduce a la definición de una relación de precedencia, a menudo es indispensable incluir la propiedad de reflexividad de modo que la terminología y resultados en conexión con las relaciones de orden parcial puedan aplicarse. El prerrequisito estructura de los cursos conduce de manera natural a la definición de una relación de precedencia. Sin embargo, al agregar la propiedad de reflexividad, convertimos a *R* en una relación de orden parcial.

larga en el conjunto parcialmente ordenado A es c. Esto significa que hay c cursos que deben tomarse uno después de otro. Así, bajo ninguna circunstancia un estudiante podrá terminar los cursos requeridos en menos de c semestres. Supongamos que el tamaño de la anticadena más larga en el conjunto parcialmente ordenado A es d. Esto significa que un estudiante podrá tomar a lo más d cursos requeridos en un semestre cualquiera. ❐

Presentamos como otro ejemplo un teorema que muestra una estrecha relación entre cadenas y anticadenas.'"

Teorema 4.1

Sea $(P, \leq)$ un conjunto parcialmente ordenado. Supongamos que la longitud de la cadena más grande en P es n. Entonces los elementos en P pueden ser particionados en n anticadenas disjuntas.

DEMOSTRACIÓN Demostraremos el teorema por inducción sobre n.

Base de la inducción. Para $n = 1$, no hay dos elementos en P que estén relacionados. Es claro que ellos constituyen una anticadena.

Paso de inducción. Supongamos que el teorema es válido cuando la longitud de las cadenas más largas en un conjunto parcialmente ordenado es $n - 1$. Sea P un conjunto parcialmente ordenado donde la longitud de sus cadenas más largas es igual a n. Sea M el conjunto de elementos maximales de P. Es claro, Mes una anticadena no vacía. Ahora consideremos el conjunto parcialmente ordenado $(P - M, \leq)$. Ya que no existe una cadena de longitud n en $P - M$, la longitud de las cadenas más largas es a lo más $n - 1$. Por otro lado, si la longitud de las cadenas más largas en $P - M$ es menor que $n - 1$, M debe contener dos o más elementos que son miembros de la misma cadena, lo cual es imposible. Por tanto, concluimos que la longitud de la cadena más larga en $P - M$ es $n - 1$ y, de acuerdo con la hipótesis de inducción, $P - M$ puede ser particionado en $n - 1$ anticadenas disjuntas. Así, P puede ser particionado en n anticadenas disjuntas. □

Por ejemplo, para el conjunto parcialmente ordenado P en la figura 4.18a, ya que la longitud de la cadena más larga en P es 4, los elementos en P pueden ser particionados en cuatro anticadenas disjuntas. Las figuras 4.186 y 4.18c muestran dos de tales particiones. Si aplicamos el teorema 4.1 al ejemplo 4.1 podemos estar seguros que si la longitud de la cadena más larga en A es c, entonces, en efecto un estudiante podrá completar todos los requerimientos en c semestres. Una consecuencia directa del teorema 4.1 puede establecerse como:

Corolario 4.1.1

Sea $(P, \leq)$ un conjunto parcialmente ordenado que contiene $mn + 1$ elementos. Existe una anticadena que contiene $m + 1$ elementos, o existe una cadena de longitud $n + 1$ en P.

DEMOSTRACIÓN Supongamos que la longitud de las cadenas más largas en P es n. De acuerdo con el teorema 4.1, P puede ser particionado en n anticadenas disjuntas. Si cada una de estas anticadenas consiste en m o menos elementos, el número total de elementos en P es a lo más nm, lo cual contradice la suposición del corolario.

---

† El teorema 4.1 es un dual del conocido *teorema de Dilworth*, el cual establece que si el tamaño de las anticadenas más largas en P es n, entonces los elementos de P pueden ser particionados en n cadenas disjuntas. Para una demostración del teorema de Dilworth, consulte a Mirsky [6].

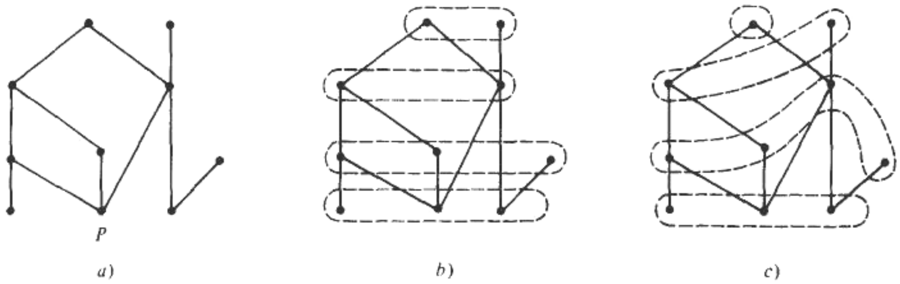


Figura 4.18

¡Ejemplo 4.21

Mostrar que de entre $ab + 1$ ratones blancos existe una sucesión de $a + 1$ ratones, cada uno descendiente del siguiente, o bien un grupo de $b + 1$ ratones, ninguno de los cuales es descendiente de algún otro. Ordenemos los ratones de acuerdo con la relación de descendencia. Es obvio que la relación de ordenamiento es una relación de orden parcial. Si existe una anticadena de tamaño $b + 1$ o mayor en este conjunto parcialmente ordenado, existe un grupo de $b + 1$ o más ratones, donde ninguno de ellos es descendiente de algún otro. Por otro lado, si existe una cadena de longitud $a + 1$ o mayor, existe una sucesión de $a + 1$ o más ratones, y cada uno es descendiente del siguiente.

Ejemplo 4.3

Un punto (x, y) en el primer cuadrante del plano xy define un rectángulo con los puntos $(0, 0)$, $(x, 0)$, $(0,y)$, (x,y) como sus vértices (véase la figura 4.19). Deseamos mostrar que para los rectángulos definidos por cinco puntos distintos cualesquiera en el primer cuadrante existen ya sea tres rectángulos $R_{i_1}, R_{i_2}, R_{i_3}$ tales que R_{i_1} y R_{i_2} están dentro de R_{i_3}, y R_{i_1} está dentro de R_{i_2}, o bien existen tres rectángulos tales que ninguno está dentro de otro. Sea $P = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)\}$ el conjunto de cinco puntos en el primer cuadrante.

Definimos una relación de orden parcial < sobre P tal que $(x_i, y_i) \leq (x_j, y_j)$ si y sólo si $x_i \leqq x_j$ y $y_i \leqq y_j$. De acuerdo con el corolario 4.1.1, existe una cadena de longitud 3 en P, o existe una anticadena de tamaño 3 en P. Es evidente que una cadena de longitud 3 corresponde a un conjunto de tres rectángulos $R_{i_1}, R_{i_2}, R_{i_3}$, tales que R_{i_1} y R_{i_2} están dentro

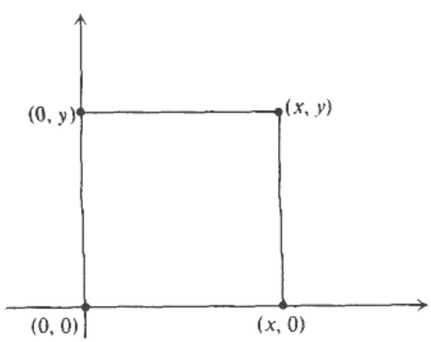


Figura 4.19

de R_{i_3}, y R_{i_1} está dentro de R_{i_2}. Por otro lado, una anticadena de tamaño 3 corresponde a un conjunto de tres rectángulos tales que ninguno se encuentra dentro de algún otro. ❐

4.7 UN PROBLEMA DE PROGRAMACIÓN DE TAREAS

Consideremos el problema de programar la realización de un conjunto de tareas en un sistema de cómputo de multiprocesador, el cual tiene un conjunto de procesadores idénticos (el problema también se puede establecer como la realización de un itinerario para un cierto número de trabajadores con el objeto de terminar un conjunto dado de tareas). Sean P_1, P_2, ..., P_n los n procesadores idénticos en un sistema de cómputo de multiprocesador. Sea $\mathcal{T} = \{T_1, T_2, \ldots, T_r\}$ el conjunto de tareas a ser ejecutadas en el sistema de cómputo. Supongamos que la ejecución de una tarea ocupa uno y sólo un procesador. Además, ya que los procesadores son idénticos, una tarea puede ser ejecutada sobre uno cualquiera de los procesadores. Sea $\mu(T_i)$ el *tiempo de ejecución* de la tarea T_i, esto es, la cantidad de tiempo que lleva ejecutar T_i en un procesador. También existe una relación de orden parcial $\leq$ especificada sobre $\mathcal{T}$ tal que para $T_i \neq T_j$, $T_i \leq T_j$ si y sólo si la ejecución de la tarea T_j no puede comenzar hasta que la ejecución de la tarea T_i haya sido finalizada (se dice que T_i, es un predecesor de T_j, y que T_j es un sucesor de T_i). Un conjunto parcialmente ordenado de tareas puede describirse mediante gráficas, como se ilustra en la figura 4.20*a*, donde el tiempo de ejecución de cada tarea se escribió junto al nombre de la tarea. Puede darse una interpretación obvia para nuestro modelo de un conjunto de tareas. Consideremos las tareas T_1, T_2, . . . , T_r como subprogramas de un programa más grande. Entonces $T_i \leq T_j$ podría significar que el subprograma T_j utiliza algunos de los datos generados por el subprograma T_i, de manera que la ejecución de T_j debe esperar la finalización de T_i. Por ejemplo, si una red de computadoras es usada en una misión espacial, la tarea T_i podría ser un subprograma que determina el curso de la nave espacial, y la tarea T_j podría ser un subprograma que estima el consumo total de combustible para el ajuste de un curso de vuelo promedio. Es claro que debemos finalizar T_i antes de ejecutar T_j.†

Por *programación* de un conjunto de tareas en un sistema de cómputo de multiprocesador, entendemos especificar para cada tarea T_i tanto el intervalo de tiempo dentro del cual ésta será ejecutada, como el procesador P_k en el cual tendrá lugar la ejecución (sin pérdida de generalidad, suponemos que la ejecución del conjunto comienza al tiempo $t = 0$). Una forma explícita para describir una programación es un *diagrama de tiempos*. Por ejemplo, el diagrama de tiempos para una programación que ejecuta el conjunto de tareas de la figura 4.20*a* en un sistema de cómputo de tres procesadores se muestra en la figura 4.206, donde ϕ_1, ϕ_2, ϕ_3, ϕ_4 denotan los periodos dentro de los cuales un procesador se deja inactivo. Para una programación determinada, un *periodo inactivo* de un procesador se define como el intervalo de tiempo dentro del cual el procesador no ejecuta una tarea. Usamos ϕ_1, ϕ_2, ... para denotar periodos inactivos de los procesadores, y $\mu(\phi_1)$, $\mu(\phi_2)$ . . . para denotar la

† Fuera del contexto de la computación, también podemos tener interpretaciones como "uno no puede ponerse los zapatos antes de ponerse los calcetines", o "la máquina no puede ensamblarse hasta que todas sus subpartes hayan sido construidas".

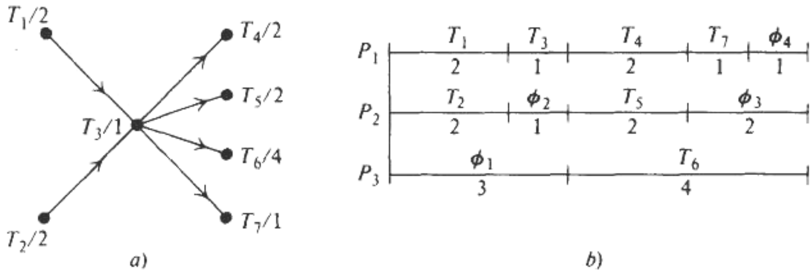


Figura 4.20

duración de los periodos inactivos. Observemos que en una programación dada, un procesador podría dejarse inactivo ya sea porque no hay una tarea ejecutable* en tal tiempo, o porque ésta es una opción intencional. Es obvio que nunca será necesario ni útil en una programación dada dejar todos los procesadores inactivos al mismo tiempo. Por otro lado, como veremos adelante, podría ser útil dejar algunos de los procesadores inactivos, aunque hay tareas que son ejecutables en ese momento.

El *tiempo de ejecución* total de una programación es el tiempo total requerido para finalizar la ejecución de todas las tareas de acuerdo con la programación. Claro que es deseable obtener una programación que tenga el mínimo tiempo de ejecución total. Desafortunadamente, no existe un procedimiento conocido (que haga más corto el método exhaustivo de tanteos) para construir programaciones con un mínimo tiempo de ejecución total. En consecuencia, otra manera de afrontar el problema es buscar programaciones que sean buenas pero no necesariamente las mejores posibles. De hecho, existe una manera muy simple e intuitiva para programar un conjunto dado de tareas, a saber, nunca dejar intencionalmente un procesador inactivo. Es decir, un procesador se deja inactivo por un periodo de tiempo sólo si ninguna tarea es ejecutable durante tal periodo. Por el contrario, siempre que un procesador esté disponible en un tiempo cualquiera, ejecutaremos en este procesador cualquier tarea que sea ejecutable en tal momento. Por ejemplo, la programación en la figura 4.20*b* para el conjunto de tareas de la figura 4.20*a* fue obtenida de esta manera. Al inicio, ya que sólo T_1 y T_2 son ejecutables en tal tiempo, éstas son ejecutadas en los procesadores P_1 y P_2, respectivamente. La finalización de T_1 y T_2 conduce a la ejecución de T_3. Después que la ejecución de T_3 ha terminado, T_4, T_5, T_6 *y* T_7 se convierten todos en ejecutables. Una opción arbitraria para ejecutar T_4, T_5, T_6 en P_1, P_2, P_3, respectivamente, origina la programación de la figura 4.20*b*. Nos apresuramos a señalar que esta manera simple de programar tareas no siempre nos da una programación con el mínimo de tiempo de ejecución total. En realidad, la figura 4.21*b* muestra una programación para el conjunto de tareas en la figura 4.2 la en la cual entre $t = 9$ y $t = 10$, el procesador P_2 se deja inactivo aunque T_3 es ejecutable en $t = 9$. El lector puede convencerse de que la programación en la figura 4.21*b* es mejor que cualquier otra en la que los procesadores no se dejan inactivos de modo intencional.

Queremos determinar qué tan bueno o malo es un procedimiento de programación simple como el anterior y llegamos a un resultado inesperado:

† Se dice que una tarea es *ejecutable* en un cierto instante si la ejecución de sus predecesores ha sido finalizada en dicho tiempo.

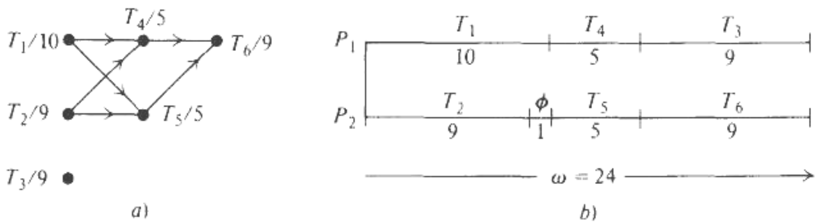


Figura 4.21

Teorema 4.2 Para un conjunto de tareas dado, sea ω el tiempo de ejecución total cuando las tareas son ejecutadas de acuerdo con una programación que no contiene periodos intencionalmente inactivos, y sea ω_0 el mínimo posible para el tiempo de ejecución total. Entonces,

$$\frac{\omega}{\omega_0} \leqq 2 - \frac{1}{n}$$

donde n es el número de procesadores en el sistema de cómputo. Además, la cota es la mejor posible.

DEMOSTRACIÓN Para simplificar la presentación, demostraremos el resultado para $n = 2$. La demostración del caso general es análoga. Consideremos la programación que no contiene periodos intencionalmente inactivos, como se ilustra en la figura 4.22. Observemos, primero, que el término de un periodo inactivo en un procesador coincide con la finalización de la ejecución de una tarea en otro procesador (de otra manera, el periodo inactivo no podría ser determinado). Sea ϕ_i un periodo inactivo para un procesador. Diremos que una tarea T_{ij} traslapará a ϕ_i si el tiempo de ejecución de T_{ij} en el otro procesador traslapa a ϕ_i. Por ejemplo, en la figura 4.22, T_{11}, T_{12}, T_{13} traslapan a ϕ_1. Sean $T_{i1}, T_{i2}, T_{i3}, \ldots, T_{il}$ las tareas que traslapan a ϕ_i. Requerimos que

$$T_{i1} \leq T_{i2} \leq T_{i3} \leq \cdots T_{il}$$

Si éste no fuera el caso, las tareas $T_{i1}, T_{i2}, T_{i3}, \ldots, T_{il}$ podrían no tener que ejecutarse en forma secuencial en P_2, y alguna de ellas podría ser ejecutada en P, en lugar de

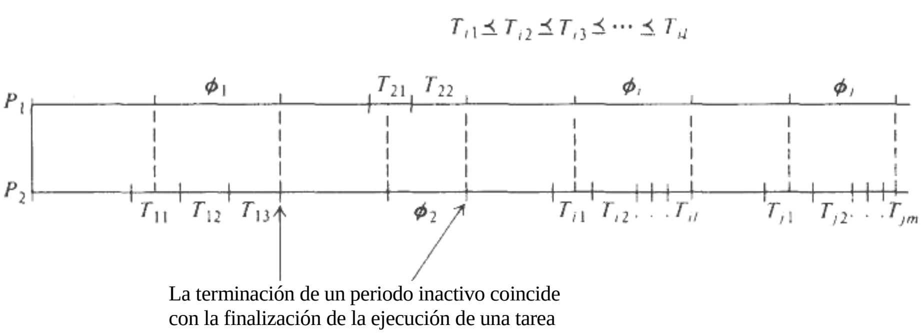


Figura 4.22

algunas porciones del periodo inactivo ϕ_i en P_1. De modo similar, si ϕ_j es otro periodo inactivo, y $T_{j1}, T_{j2}, T_{j3}, \ldots, T_{jm}$ son las tareas que se traslapan con ϕ_j, entonces repitiendo el mismo argumento tenemos

$$T_{j1} \leq T_{j2} \leq T_{j3} \leq \cdots \leq T_{jm}$$

Es obvio que cualquier tarea ejecutada después de la finalización de T_{il} debe ser un sucesor de T_{il} (o de lo contrario ésta debería ejecutarse en ϕ_i). En consecuencia, existe un subconjunto de tareas $\mathscr{C}$ tal que:

1. $\mathscr{C}$ es una cadena.
2. $\sum_{T_k \in \mathscr{C}} \mu(T_k) \geqq \sum_{\phi_i \in \Phi} \mu(\phi_i)$,

donde Φ es el conjunto de todos los periodos inactivos en la programación. Observamos que

$$\begin{aligned} \omega &= \frac{1}{2}\left[\sum_{T_j \in T} \mu(T_j) + \sum_{\phi_i \in \Phi} \mu(\phi_i)\right] \\ &\leqq \frac{1}{2}\left[\sum_{T_j \in T} \mu(T_j) + \sum_{T_k \in \mathscr{C}} \mu(T_k)\right] \end{aligned} \tag{4.1}$$

En efecto,

$$\omega_0 \geqq \frac{1}{2} \sum_{T_j \in T} \mu(T_j)$$

y

$$\omega_0 \geqq \sum_{T_k \in \mathscr{C}} \mu(T_k)$$

Así, (4.1) se convierte en

$$\omega \leqq \omega_0 + \tfrac{1}{2}\omega_0$$

o

$$\frac{\omega}{\omega_0} \leqq \frac{3}{2}$$

Puede demostrarse, mediante el ejemplo de la figura 4.23, que esta cota es la mejor posible; la figura 4.23*b* muestra dos programaciones para el conjunto de tareas en la figura 4.23*a*.

❐

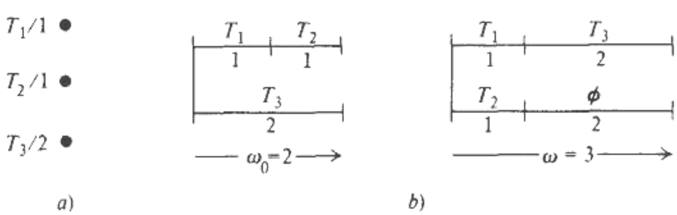


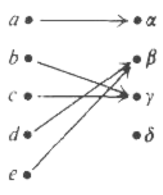
Figura 4.23

Es interesante observar que, de acuerdo con el resultado que acabamos de demostrar, la programación obtenida al seguir la regla simple de no dejar intencionalmente un procesador inactivo nunca es peor que la mejor programación posible por más de 50% en un sistema de cómputo de dos procesadores, y por más de 100% en un sistema de cómputo de «-procesadores. Es claro que este resultado es un gran apoyo a la idea de emplear procedimientos simples para obtener buenos, pero no necesariamente los mejores resultados posibles, en muchos problemas de optimización.

4.8 FUNCIONES Y EL PRINCIPIO DEL PALOMAR

Se dice que una relación binaria *R* de *A* a *B* es una *función* si para cualquier elemento *a* en *A*, existe un único elemento *b* en *B* tal que (a, *b)* está en *R.* Para una función *R* de *A* a *B,* en lugar de escribir *(a, b)* e *R,* usaremos la notación *R(a)* = *b,* donde *b* es llamado la *imagen* de *a.* El conjunto *A* es llamado el *dominio* de la función *R,* y el conjunto *B* es llamado el *rango* de la función *R.* La noción de una función no es sino una formalización de la noción de asociar o asignar un elemento en el rango a cada uno de los elementos en el dominio. Por ejemplo, *sea A* un conjunto de casas y *B* es un conjunto de colores, entonces una función de *A* a *B* es una asignación de colores para pintar las casas. A menudo una función puede ser representada en forma gráfica. La figura 4.24*a* muestra una función *R* de *A* = *{a, b, c, d, e}* a $B = \{\alpha, \beta, \gamma, \delta\}$. Si seguimos la convención de representar una relación binaria en forma tabular, podemos representar la función de la figura 4.24*a* como la de la figura 4.24*b.* No obstante, una forma tabular más conveniente para representar funciones es la mostrada en la figura 4.24*c*, donde la columna izquierda contiene todos los elementos en el dominio y la columna derecha contiene sus imágenes correspondientes.

Se dice que una función de *A* a *B* es una función *sobre* si cualquier elemento de *B* es la imagen de uno o más elementos de *A.* La figura 4.25*a* muestra un ejemplo de una función sobre. Se dice que una función de *A* a *B* es una función *uno a uno* si no hay dos elementos de *A* que tengan la misma imagen. La figura 4.256 muestra un ejemplo de una función uno a uno. Se dice que una función de *A* a *B* es una función *uno a uno sobre* si ésta es tanto una



a)

	α	β	γ	δ
a	√			
b			√	
c			√	
d		√		
e		√		

b)

	R
a	α
b	γ
c	γ
d	β
e	β

c)

Figura 4.24

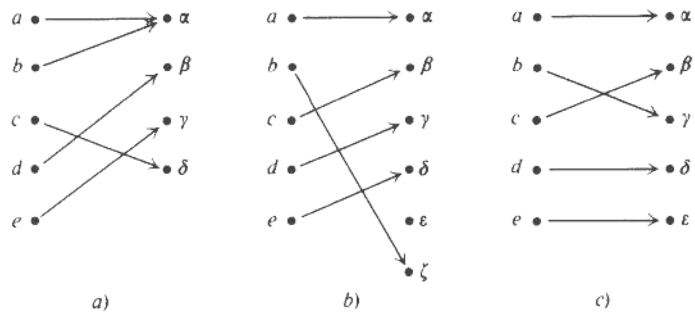


Figura 4.25

función sobre como una función uno a uno.† La figura 4.25c muestra una función uno a uno sobre. *Sea A* un conjunto de trabajadores y B_1 B_2 y B_3 conjuntos de trabajos. Una función sobre de A a B_1 es una asignación de los trabajadores a los trabajos de manera que cada trabajo tiene al menos un trabajador asignado; una función uno a uno de A a B_2 es una asignación tal que no hay dos trabajadores que tengan el mismo trabajo, y una función uno a uno sobre de A a B_3 es una asignación tal que cada trabajo tiene un trabajador asignado, y no hay dos trabajadores asignados al mismo trabajo.

En la literatura, una función sobre también se conoce como *sobreyectiva*, una función uno a uno como *inyectiva*, y una función uno a uno sobre como *biyectiva.*

En matemáticas una técnica de demostración conocida es la denominada como *principio del palomar*, también conocida como el *argumento de la caja de zapato* o *el principio del cajón de Dirichlet.* De manera informal el principio del palomar dice que si hay "muchas" palomas y "unos pocos" palomares, entonces debe haber algunos palomares ocupados por dos o más palomas. Formalmente, sean D y R conjuntos finitos. Si $|D| > |R|$, entonces para cualquier función f de D a R, existen $d_1, d_2 \in D$ tales que $f(d_1) = f(d_2)$. Algunas aplicaciones triviales del principio del palomar son: de entre 13 personas, hay al menos 2 de ellas que nacieron en el mismo mes. Aquí las 13 personas son las palomas, y los 12 meses son los palomares. De igual manera, si 11 zapatos son seleccionados de 10 pares de zapatos, debe haber un par de zapatos que sean el par entre la selección. Aquí los 11 zapatos son las palomas y los 10 pares de zapatos son los palomares. El principio del palomar puede establecerse en una forma más general: para cualquier función/de D a R, existen / elementos $d_x, d_2, \ldots, d_i$ en D, $i = [|D|/|R|]$ tales que $f(d_1) = f\{d_2) = \ldots = f\{d$‡ (véase problema 4.34).

En los siguientes ejemplos, exhortamos al lector a observar cómo es aplicado el principio del palomar, ya que no haremos una declaración explícita cada vez que lo usemos.

---

† Recuerde que introdujimos la noción de correspondencia uno a uno entre los elementos en dos conjuntos. Formalmente, decimos que hay una correspondencia uno a uno entre los elementos de dos conjuntos si existe una función biyectiva de un conjunto al otro.

‡ Usamos F x 1 para denotar el entero más pequeño no menor que *x*.

Ejemplo 4.4

Una jugadora de ajedrez desea prepararse para un encuentro de campeonato realizando algunas partidas de práctica durante 77 días. Ella desea jugar al menos una partida por día pero no más de 132 partidos en total. Ahora mostraremos que sin importar cómo programe ella las partidas habrá un periodo de días consecutivos dentro del cual ella jugará *exactamente* 21 partidos. Sea a_i el número total de partidos que ella juega a lo largo del i-ésimo día. Es obvio, la sucesión $a_1, a_2, \ldots, a_{77}$ es una sucesión monótona creciente, con $a_1 \geqq 1$ y $a_{77} \leqq 132$. Calculamos la sucesión $a_1 + 21$, $a_2 + 21$,..., $a_{77} + 21$, la cual de nuevo es monótona creciente con $a_{77} + 21 \leqq 153$. Puesto que los valores de los 154 números $a_1, a_2, \ldots, a_{77}, a_1 + 21, \ldots, a_{77} + 21$ van desde 1 hasta 153, dos de ellos deben ser el mismo. Además, debido a que tanto la sucesión $a_1, a_2, \ldots, a_{77}$ como la sucesión $a_1 + 21$, $a_2 + 21$,..., $a_{77} + 21$ son monótonas crecientes, tenemos que $a_i = a_j + 21$ para algún a_i y algún a_j.

□

Ejemplo 4.5

Deseamos demostrar que en una sucesión de $n^2 + 1$ enteros distintos, existe ya sea una subsucesión creciente de longitud $n + 1$ o una subsucesión decreciente de longitud $n + 1$. Sea $a_1, a_2, \ldots, a_{n^2+1}$ sucesión de enteros. Etiquetamos al entero a_k con un par ordenado $(x_k\ y_k)$, donde x_k es la longitud de la subsucesión creciente más larga a partir de a_k, y y_k es la longitud de la subsucesión decreciente más larga a partir de a_k. Supongamos que no existe una subsucesión creciente o una subsucesión decreciente de longitud $n + 1$ en la sucesión $a_1, a_2, \ldots, a_{n^2+1}$. Esto es, los valores de x_k y y_k están entre 1 y n para $k = 1, 2, \ldots, n^2 + 1$. Con sólo n^1 pares ordenados distintos como posibles etiquetas para los $n^2 + 1$ enteros, deben existir a_i y a_j en la sucesión que estén etiquetados con el mismo par ordenado. No obstante, esto es imposible porque si $a_i < a_j$, debemos tener que $x_i > x_j$, y si $a_i > a_j$, debemos tener $y_i > y_j$. Por tanto, concluimos que existe ya sea una subsucesión creciente, o una subsucesión decreciente de longitud $n + 1$ en la sucesión $a_1, a_2, \ldots, a_{n^2+1}$.†

□

Ejemplo 4.6

Mostraremos que de entre seis personas, hay tres personas que son amigos o hay tres personas que son desconocidas por completo entre sí. Sea A una persona en el grupo. De acuerdo con el principio del palomar, hay tres (o más) personas que son amigos de A o hay tres (o más) personas que son extraños para A. Consideremos el primer caso solamente, ya que el último caso puede resolverse mediante un argumento similar. Si B, C, D son los amigos de A, y si cualesquiera dos de B, C, D se conocen uno al otro, entonces estos dos, con A, forman una tercia de conocidos. Por otro lado, si no hay dos personas de entre B, C, D que se conozcan entre sí, entonces hay tres personas que son completamente desconocidas entre sí.

Ejemplo 4.7

Una casa de huéspedes tiene 90 habitaciones y 100 huéspedes. Las llaves son distribuidas a los huéspedes de manera que cualesquiera 90 huéspedes pueden tener acceso a las

† Véase el problema 4.41 para un argumento ligeramente diferente. También el problema 4.42.

90 habitaciones, en el sentido de que cada huésped tendrá una llave de una habitación desocupada (en otras palabras, podemos asignar las 90 habitaciones a cualesquiera 90 huéspedes de manera que cada huésped tenga una llave para la habitación a la que ha sido asignado). Deseamos usar un esquema que minimice el número total de llaves. A pesar de que el siguiente esquema parece ser derrochador, a final de cuentas es un esquema minimizador. A 90 de los huéspedes les será dada una llave, de manera que todos ellos tendrán acceso a las 90 habitaciones. A cada uno de los restantes 10 huéspedes les serán dadas 90 llaves, una llave por cada una de las 90 habitaciones. Es claro que este esquema, el cual requiere un total de 990 llaves, funciona. Para mostrar que éste es un esquema minimal, observamos que, si 989 o menos llaves fueron distribuidas, existe una habitación que tiene a lo más 10 llaves sin repetir. Es obvio que el esquema falla cuando ninguno de los poseedores de llaves de esta habitación está entre los 90 huéspedes.

Ejemplo 4.8

Las circunferencias de dos discos concéntricos están divididas en 200 secciones cada una, como se muestra en la figura 4.26. Para el disco exterior, 100 de las secciones están pintadas de rojo y 100 de azul. Para el disco interior las secciones están pintadas de rojo y azul en forma arbitraria. Mostraremos que es posible alinear los dos discos de manera que 100 o más de las secciones sobre el disco interno tengan sus colores acoplados con las secciones correspondientes sobre el disco exterior. Mantendremos el disco exterior fijo y rotaremos el disco interior a través de los 200 posibles alineamientos. Por cada alineamiento, contamos el número de acoplamientos de color. Observemos que el conteo para los 200 posibles alineamientos es 20 000, porque cada una de las 200 secciones sobre el disco interno se acoplará a su correspondiente sección en el disco exterior exactamente en 100 de los alineamientos. Luego, debe haber un alineamiento en el cual hay 100 o más acoplamientos.

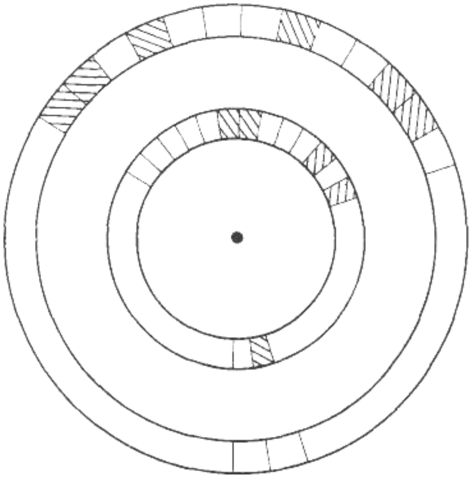


Figura 4.26

4.9 NOTAS Y REFERENCIAS

Para un análisis más extenso sobre sistemas para el manejo de bases de datos, consulte Codd [1] y Date [3]. Consulte Mirsky [6] para el teorema de Dilworth y sus aplicaciones. Las nociones de partición y pares de particiones son muy útiles en el estudio de las propiedades estructurales de las máquinas de estado finito. Consulte, por ejemplo, el capítulo 12 de Kohavi [5]. El teorema 4.2 se debe a Graham [4], Coffman [2] cubre múltiples aspectos del problema de programación de tareas introducido en la sección 4.7. Consulte el capítulo 4 de Ryser [7] para el teorema de Ramsey, el cual es una generalización no trivial del principio del palomar.

1. Codd, E. R: "A Relational Model of Data for Large Shared Data Bankcs", *Communications of ACM*, **13:** 377-387 (1970).
2. Coffman, E. G. Jr. (ed): *Computer and Job-Shop Scheduling Theory,* John Wiley & Sons, Nueva York, 1976.
3. Date, C. J.: *An Introduction to Database Systems*, 3ª ed., Addison-Wesley Publishing Company, Reading, Mass., 1981.
4. Graham, R. L.: "Bounds for Certain Multiprocessing *Anomalies", Bell System Technical Journal,* 45: 1563-1581 (1966).
5. Kohavi, Z.: *Switching and Finite Automata Theory, 2"* ed., McGraw-Hill Book Company, Nueva York, 1978.
6. Mirsky, L.: *Transversal Theory,* Academic Press, Nueva York, 1971.
7. Ryser, H. J.: *Combinatorial Mathematics,* publicado por la Mathematical Association of America, distribuido por John Wiley & Sons, Nueva York, 1963.

PROBLEMAS

4.1 Si *A* denota el conjunto de camisas y *B* denota el conjunto de tirantes que un hombre posee, ¿qué posible interpretación puede darse al producto cartesiano *A* x 5?, ¿a la relación binaria de *A* a *Bl*

4.2 Sea $A = \{1,2\}$. Construya el conjunto $\mathscr{P}(A) \times A$.

4.3 *a)* Dado que $A \subseteq C$ y $B \subseteq D$, demuestre que $A \times B \subseteq C \times D$.

b) Dado que $A \times B \subseteq C \times D$, ¿se sigue de esto necesariamente que $A \subseteq C$ y $B \subseteq D$?

4.4 *a)* *Sea A* un conjunto arbitrario. ¿Está bien definido el conjunto $A \times \varnothing$?

b) Dado que $A \times B = \varnothing$, ¿qué puede decir uno acerca de los conjuntos *A* y 5?

c) ¿Es posible que $A \subseteq A \times A$ para algún conjunto *Al*

4.5 Sean *A, B, C, D* conjuntos arbitrarios.

a) Demuestre que

$$(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$$

b) Demuestre o dé un contraejemplo de las siguientes identidades:

$$(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D)$$
$$(A - B) \times (C - D) = (A \times C) - (B \times D)$$
$$(A \oplus B) \times (C \oplus D) = (A \times C) \oplus (B \times D)$$

4.6 Sean *A, B, C* conjuntos arbitrarios.

a) Demuestre que

$$(A \cap B) \times C = (A \times C) \cap (B \times C)$$

b) Demuestre o dé un contraejemplo de las siguientes identidades:

$$(A \cup B) \times C = (A \times C) \cup (B \times C)$$
$$(A - B) \times C = (A \times C) - (B \times C)$$
$$(A \oplus B) \times C = (A \times C) \oplus (B \times C)$$

4.7 Sea A = {6:00, 6:30, 7:00, . . . , 9:30, 10:00} el conjunto de 9 periodos de media hora por la mañana, y B = {3, 12, 15, 17} el conjunto de los cuatro canales locales de televisión, y R_1 y R_2 las dos relaciones binarias *de A* a 5. ¿Qué posibles interpretaciones pueden darse a las relaciones binarias R_1, R_2, $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 \oplus R_2$ y $R_1 - R_2$?

4.8 Sea I el conjunto de todos los enteros.

a) ¿Existe una manera natural de interpretar los pares ordenados en I x I como puntos geométricos en el plano?

b) Sea R_1, una relación binaria sobre I x I tal que el par ordenado (de pares ordenados) ((a, *b*), (*c*, *d*)) está en R_1 si y sólo si $a - c = b - d$. ¿Cuál es la interpretación geométrica de la relación binaria R_1?

c) Si R_2 es una relación binaria sobre I x / tal que *((a, b), (c, d))* está en R_2 si y sólo si

$\sqrt{(a-c)^2 + (b-d)^2} \leqq 10$. ¿Cuál es la interpretación geométrica de R_2? ¿De $\bar{R}_1 \cup R_2$? ¿De $R_1 \cap R_2$? ¿De $R_1 - R_2$? ¿De $R_1 \oplus R_2$?

4.9 Sea A un conjunto de trabajadores y B un conjunto de trabajos. Considere que R_1, es una relación binaria de A a B tal que *(a, b)* está en R_x si el trabajador *a* está asignado al trabajo *b* (supongamos que un trabajador puede ser asignado a más de un trabajo y que más de un trabajador puede ser asignado al mismo trabajo). Sea R_2 una relación binaria sobre A tal que (a_1 a_2) está en R_2 si a_1 y a_2 pueden convivir uno con el otro si fueran asignados al mismo trabajo. Establezca una condición en términos de R_1 R_2 *y* (posibles) relaciones binarias derivadas a partir de R_1, y R_2 de tal manera que una asignación de los trabajadores a los trabajos de acuerdo con R_1, no colocará trabajadores que no puedan convivir uno con el otro para realizar el mismo trabajo.

4.10 Sea A un conjunto de libros.

a) Sea 7?, una relación binaria sobre A tal que *(a, b)* está en R_1 si el libro *a* cuesta más y contiene menos hojas que el libro *b*. En general, ¿es R_1 reflexiva?, ¿simétrica?, ¿antisimétrica?, ¿transitiva?

b) Sea R_2 una relación binaria sobre A tal que *(a, b)* está en R_2 si el libro *a* cuesta más o contiene menos hojas que el libro *b*. En general, ¿es R_2 reflexiva?, ¿simétrica?, ¿antisimétrica?, ¿transitiva?

4.11 *a)* Sea R una relación binaria sobre el conjunto de todos los enteros positivos tal que

$$R = \{(a, b) \mid a - b \text{ es un entero positivo impar}\}$$

¿Es R reflexiva?, ¿simétrica?, ¿antisimétrica?, ¿transitiva?, ¿una relación de equivalencia?, ¿una relación de orden parcial?

b) Repita el inciso (*a*) si $R = \{(a, b) \setminus a = b^2\}$

4.12 Sea P el conjunto de toda la humanidad. Sea R una relación binaria sobre P tal que *{a, b)* está en R si *a* es un hermano de *b* (considere sólo como hijos a los de la misma pareja padre-madre). ¿Es R reflexiva?, ¿simétrica?, ¿antisimétrica?, ¿transitiva?, ¿una relación de equivalencia?, ¿una relación de orden parcial?

4.13 Sea R una relación binaria sobre el conjunto de todas las cadenas de ceros y unos tal que R = {*(a, b)* | *a y b* son cadenas que tienen el mismo número de ceros}. ¿Es R reflexiva?, ¿simétrica?, ¿antisimétrica?, ¿transitiva?, ¿una relación de equivalencia?, ¿una relación de orden parcial?

4.14 Utilice el hecho de que "mejor que" es una relación binaria transitiva para demostrar la afirmación "un emparedado de jamón es mejor que la felicidad eterna".

Sugerencia: Nada es mejor que la felicidad eterna.

4.15 Sea A un conjunto con 10 elementos distintos.

a) ¿Cuántas relaciones binarias diferentes hay sobre A?

b) ¿Cuántas de ellas son reflexivas?
c) ¿Cuántas de ellas son simétricas?
d) ¿Cuántas de ellas son reflexivas y simétricas?
e) ¿Cuántas de ellas son relaciones de orden total?

4.16 Sea R una relación simétrica y transitiva sobre un conjunto A. Demuestre que si para todo a en A existe un b en A tal que (a, b) está en R, entonces R es una relación de equivalencia.

4.17 Sea R una relación transitiva y reflexiva sobre A. Sea *Tuna* relación sobre A tal que (a, b) está en T si y sólo si tanto (a, b) como (b, a) están en R. Demuestre que T es una relación de equivalencia.

4.18 Sea R una relación binaria. Sea $S = \{(a, b) \mid (a, c) \in R \text{ y } (c, b) \in R \text{ para algún } c\}$. . Demuestre que si R es una relación de equivalencia, entonces S también es una relación de equivalencia.

4.19 Sea R una relación reflexiva sobre un conjunto A. Demuestre que i? es una relación de equivalencia si y sólo si (a, b) y (a, c) están en R, y que entonces $(b$, c) está en R.

4.20 Una relación binaria sobre un conjunto que es reflexiva y simétrica se llama *relación compatible.*

a) Sea A un conjunto de personas y R una relación binaria sobre A tal que (a, b) está en R si a es un amigo de b. Demuestre que R es una relación compatible.
b) Sea A un conjunto de palabras inglesas y R una relación binaria sobre A tal que dos palabras en A están relacionadas si tienen una o más letras en común. Demuestre que R es una relación compatible.
c) Dé más ejemplos de relaciones compatibles.
d) Sean R_j y R_2 dos relaciones compatibles sobre A.una¿Es $R_1 \cap R_2$ relación compatible? ¿Es $R_1 \cup R_2$ una relación compatible?
e) Sea A un conjunto. Una *cubierta* de A es un conjunto de subconjuntos no vacíos de A, $\{A_1$ $A_2, \ldots, A_k]$, tal que la unión de los A_j es igual a A. Sugiera una manera para definir una relación compatible sobre A a partir de una cubierta de A. Dé una interpretación de la noción de cubierta en términos del ejemplo en el inciso *a).*
f) Sugiera una manera para definir una cubierta de A a partir de una relación compatible sobre A. Sugiera una forma de definir la unicidad de la cubierta de A.

4.21 *a)* Demuestre que la cerradura transitiva de una relación simétrica es simétrica.
b) ¿Es la cerradura transitiva de una relación antisimétrica siempre antisimétrica?
c) Demuestre que la cerradura transitiva de una relación compatible (véase problema 4.20) es una relación de equivalencia.

4.22 Sea R una relación binaria de A a B. La *inversa* de R, denotada por R ~', es una relación binaria de B a A tal que

$$R^{-1} = \{(b, a) \mid (a, b) \in R\}$$

a) Sean R_J y R_2 relaciones binarias de A a B. ¿Es verdad que $(R_1 \cup R_2)^{-1} = R_1^{-1} \cup R_2^{-1}$?
b) Sea R una relación binaria sobre A. Si R es reflexiva, ¿es R^{-1} necesariamente reflexiva? Si R es simétrica, ¿es R^{-1} necesariamente simétrica? Si R es transitiva ¿es R^{-1} necesariamente transitiva?

4.23 Sea R una relación binaria sobre A, y sean $R_1, R_2, \ldots, R_i, \ldots$ las extensiones transitivas sucesivas de R como se definieron en la sección 4.3. Demuestre por inducción que si (a, b) está en R_i, existen n elementos en A, $n \leqq 2^i - 1, x_1, x_2, \ldots, x_n$ tales que $(a, x), (x_1, x_2), \ldots, (x_{n-1}, x_n), (x_n, b)$ están todas en R.

4.24 *a)* Sea R una relación de equivalencia sobre un conjunto A. Sea $\{A_1, A_2, \ldots, A_k\}$ un conjunto de subconjuntos de A tales que $A_i \not\subseteq A_j$ para $i \neq j$ y tales que a y b están contenidos en uno de los subconjuntos si y sólo si el par ordenado (a, b) está en R. Demuestre que $\{A_1, A_2, \ldots, A_k\}$ es una partición de A.
b) Sea $\{A_1\ A_2, \ldots, A_k\}$ una partición del conjunto A . Definimos una relación binaria R sobre A tal que un par ordenado (a, b) está en R si y sólo si a y b están en el mismo bloque de la partición. Demuestre que R es una relación de equivalencia.

4.25 Supongamos que S y T son dos conjuntos y f es una función de S a T. Sea R_1 una relación de equivalencia sobre T. Sea R_2 una relación binaria sobre 5 tal que $(x, y) \in R_2$ si y sólo si $(f(x), f(y)) \in R_1$. Demuestre que R_2 también es una relación de equivalencia.

4.26 Sea A un conjunto y/una función de A a A. Diremos que una partición *n de A* tiene *Impropiedad de sustitución* con respecto a f si para dos elementos cualesquiera a y b, que están juntos en un bloque de π, los dos elementos $f(a)$ y $f(b)$ también están juntos en un bloque de π. Sea $A = \{1,2,3,4, 5, 6\}$. Sea f una función de A a A tal que $f(1) = 3, f(2) = 3, f(3) = 2, f(4) = 5, f(5) = 4, f(6) = 4$.

a) ¿Tiene $\pi_1 = \{\overline{123}, \overline{456}\}$ la propiedad de sustitución con respecto a/? ¿Qué sucede acerca de $\pi_2 = \{\overline{16}, \overline{25}, \overline{34}\}$?, ¿y de $\pi_3 = \{\overline{12}, \overline{34}, \overline{56}\}$?

b) Sea A el conjunto de los enteros y π una partición de A en números pares e impares. Sea $f(a) = a + 1$ para todo a en A. ¿Tiene π la propiedad de sustitución con respecto a f? Sea

$$g(a) = \begin{cases} \dfrac{a}{2} & \text{si } a \text{ es par} \\ \dfrac{a-1}{2} & \text{si } a \text{ es impar} \end{cases}$$

¿Tiene π la propiedad de sustitución con respecto a g?

c) Sean π_1 y π_2 dos particiones que tienen la propiedad de sustitución con respecto a f. Demuestre que tanto $\pi_1 \cdot \pi_2$ como $\pi_1 + \pi_2$ tienen la propiedad de sustitución con respecto a f

4.27 Sea $(A, \leq)$ un conjunto parcialmente ordenado. Sea $<_R$ una relación binaria sobre A tal que para a y b en A, $a \leq_R b$ si y sólo si $b \leq a$.

a) Demuestre que $\leq_R$ es una relación de orden parcial.

b) Demuestre que si $(A, \leq)$ es una lattice, entonces $(A, \leq_R)$ también es una lattice.

4.28 Para un conjunto dado A, considere la relación

$$R = \{(x, y) \mid x \in \mathscr{P}(A), y \in \mathscr{P}(A) \text{ y } x \subseteq y\}$$

a) Demuestre que R es una relación de orden parcial.

b) ¿Cuál es la longitud de la cadena más larga en el conjunto parcialmente ordenado $(\mathscr{P}(A), R)$?

4.29 ¿El producto cartesiano de dos lattices siempre es una lattice? Demuestre su afirmación.

4.30 Sean P un conjunto parcialmente ordenado, y L una cadena de dos elementos. Sean Q el producto cartesiano P x L, A una anticadena *enQ,yB* el subconjunto de P más grande posible tal que B no contiene una cadena de longitud mayor a 2. Demuestre que $|A| \leqq |B|$.

4.31 El procedimiento para programar un conjunto de tareas de acuerdo con la regla de nunca dejar intencionalmente un procesador inactivo estudiado en la sección 4.7, no especificaba cómo se pueden eliminar los congestionamientos cuando varias tareas ejecutables compiten por un procesador libre. Una manera de romper congestionamientos es asignar prioridades distintas a las tareas, y programar la tarea que posee la más alta prioridad, de entre todas las tareas ejecutables, en un instante cualquiera en un procesador que está libre en dicho instante. Por ejemplo, el conjunto de tareas mostrado en la figura 4P. 1 debe ejecutarse en un sistema de cómputo con tres procesadores. La asignación de prioridades en orden decreciente es $T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9$.

a) Construya la programación correspondiente.

b) Suponga que eliminamos las flechas entre T_4 y T_5 y entre T_4 y T_6 en la figura 4P. 1. Construya la programación correspondiente.

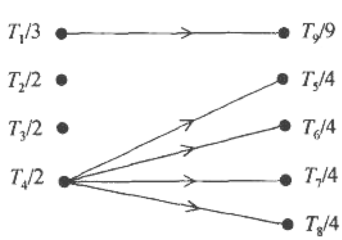


Figura 4P.1

c) Suponga que reducimos el tiempo de ejecución de cada tarea por 1. Construya la programación correspondiente.

d) Suponga que ejecutamos el conjunto de tareas en un sistema de cómputo con cuatro procesadores. Construya la programación correspondiente.

(Este problema ilustra algunas de las *anomalías* cuando las tareas son programadas de acuerdo con la regla de nunca dejar intencionalmente un procesador inactivo).

4.32 Sea f una función de A a B y g una función de B a C. Podemos definir una función h de A a C tal que para todo a en A, $h(a) = g(f(a))$. h es llamada la *composición* de f y g, y se denota por $g \circ f$

a) Determine $g \circ f$ para las funciones f y g de la figura 4P.2.

b) Determine $h \circ (g \circ f)$ y $(h \circ g) \circ f$ para las funciones f g y h de la figura 4P.2.

c) Demuestre que para funciones cualesquiera f, g y h, $h \circ (g \circ f) = (h \circ g) \circ f$,

d) Establezca una condición necesaria y suficiente sobre/ y g tal que:

1. $g \circ f$ es una función sobre;
2. $g \circ f$ es una función uno a uno;
3. $g \circ f$ es una función biyectiva.

e) Construya un ejemplo para mostrar que para dos funciones f y g de A a A, en general, $f \circ g \neq g \circ f$.

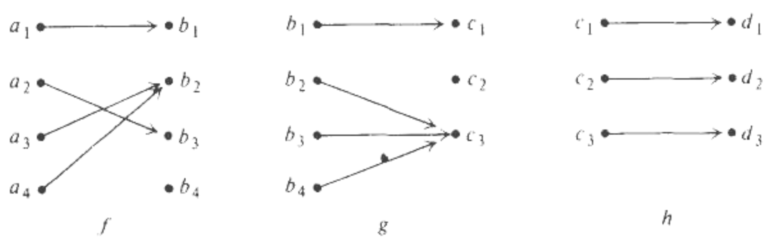


Figura 4P.2

4.33 Sean f, g y h funciones de N a N, donde N es el conjunto de números naturales de manera que

$$f(n) = n + 1$$
$$g(n) = 2n$$
$$h(n) = \begin{cases} 0 & n \text{ es par} \\ 1 & n \text{ es impar} \end{cases}$$

Determine $f \circ f, f \circ g, g \circ f, g \circ h, h \circ g, (f \circ g) \circ h$.

4.34 Sea/una función de D a R. Suponga que | D \ > | R \.

a) Sea ¡' el cociente y r el residuo cuando | D | es dividido por | R |. Demuestre que

$$\left\lceil \frac{|D|}{|R|} \right\rceil = \begin{cases} i + 1 & \text{si } r \neq 0 \\ i & \text{si } r = 0 \end{cases}$$

b) Demuestre que existen i elementos d_1, d_2,. .., d_1 en d, tales que $f(d_1) = f(d_2) = \dots = f(d_i)$.

4.35 Demuestre que entre 100 000 personas hay dos que nacieron exactamente al mismo tiempo (hora, minuto y segundo).

4.36 *a)* A partir de una baraja con 52 cartas, ¿cuántas cartas deben escogerse de manera que siempre se incluyan tres espadas en la selección?

b) Repita el inciso *a)* si la selección siempre incluye tres espadas y tres corazones.

c) Repita el inciso *a)* si dos cartas de cada palo siempre se incluyen en la selección.

4.37 *a)* A partir de 15 letras A, 20 letras B y 25 letras C, ¿cuántas letras deben escogerse de manera que siempre se incluyan 12 letras idénticas en la selección?

b) Repita el inciso *a)* si hay 15 letras A, 20 letras B, 25 letras C, 10 letras D y 8 letras E.

c) Repita el inciso *b)* si las letras D se consideran "comodines" (esto es, *i* letras D junto con 12 - *i* de cualquiera de las otras cuatro letras se consideran como 12 "letras idénticas").

4.38 ¿Cuántos números deben escogerse a partir de los enteros 10,11,12,13,..., 97,98,99 de manera que al menos un múltiplo de 3 esté incluido en la selección? ¿De manera que dos números con el mismo primer dígito sean incluidos en la selección? ¿De manera que dos números con al menos un dígito en común (por ejemplo, 12 y 52, 12 y 25) sean incluidos en la selección?

4.39 En una universidad hay 35 000 estudiantes. Cada uno de ellos toma cuatro cursos (distintos). La universidad ofrece 999 cursos diferentes. Cuando una estudiante que ha tomado un curso de matemáticas discretas se da cuenta de que el salón de clase más grande admite sólo 135 estudiantes, concluye que existe un problema. ¿Cuál es el problema?

4.40 Para jugar el juego Lotto, se compra un boleto y se seleccionan 6 de los 44 números 1, 2, 3,..., 44. En seguida, 6 de los 44 números se extraen al azar y se declaran como los números ganadores, y el gran premio es otorgado a la selección que coincide con los seis números ganadores. Suponga que se otorga un premio de consolación a una selección que no coincide con ninguno de los seis números ganadores. Para estar seguros de recibir el premio de consolación, ¿cuál es el número mínimo de boletos que uno puede comprar? ¿Cómo deberán ser escogidos los números?

4.41 En este problema demostramos el resultado del ejemplo 4.5 mediante un argumento un poco diferente. Sean $a_1, a_2, \ldots, a_{n^2+1}$ una sucesión de n^2+1 enteros distintos. Suponga que no hay una sucesión creciente de longitud $n + 1$ en la sucesión. Etiquetamos el entero a_k con una etiqueta x_k, donde x_k es la longitud de una subsucesión creciente más larga a partir de a_k, $1 \leqq k \leqq n^2+1$.

a) ¿Cuál es el significado de una cadena en el conjunto parcialmente ordenado?, ¿y de una anticadena?

b) Demuestre que estos enteros forman una subsucesión decreciente.

4.42 En este problema demostramos cómo podemos aplicar el corolario 4.1.1 para demostrar el resultado del ejemplo 4.5. Sea $a_1, a_2, \ldots, a_{n^2+1}$ una sucesión de n^2+1 enteros distintos. Definimos una relación de orden parcial $\leq$ sobre los n^2+1 pares ordenados (a_i, i) tal que $(a_i, i) \leq (a_j, j)$ si y sólo si $a_i \leqq a_j$ e $i \leqq j$.

a) ¿Cuál es el significado de una cadena en el conjunto parcialmente ordenado?, ¿y de una anticadena?

b) Aplique el corolario 4.1.1 para demostrar el resultado del ejemplo 4.5.

4.43 Demuestre que uno de *m* cualesquiera enteros consecutivos es divisible por *m*.

4.44 Demuestre que la expansión decimal de un número racional debe, después de cierto momento, ser periódica.

4.45 Un hombre caminó por 10 horas y cubrió una distancia total de 45 millas. Se sabe que caminó 6 millas en la primera hora y sólo 3 millas en la última hora. Demuestre que debió haber caminado al menos 9 millas dentro de un cierto periodo de dos horas consecutivas.

4.46 La circunferencia de una "rueda de ruleta" se divide en 36 sectores a los cuales les son asignados en alguna forma arbitraria los números 1, 2, . . . , 36. Demuestre que existen tres sectores consecutivos tales que la suma de sus números asignados es al menos 56.

4.47 Sean $x_1, x_2, \ldots, x_n$ *n* enteros arbitrarios. Demuestre que es $x_i + x_{i+1} + x_{i+2} + \cdots + x_{i+k}$ divisible por *n* para algún i y k, $i \geqq 1$, $k \geqq 0$.

4.48 A partir de los enteros 1, 2, 3,..., 200,101 de ellos se escogen al azar. Demuestre que, entre los números escogidos, existen dos tales que uno divide al otro.

4.49 Demuestre que para un entero arbitrario *N*, existe un múltiplo de *N* que contiene sólo los dígitos 0 y 7 (por ejemplo, para *N* = 3, tenemos que 259 x 3 = 777; para **N**= 4, tenemos que 1925 x 4 = 7700; para *N* = 5, tenemos que 14 x 5 = 70; para *N* = 6, tenemos que 1295 x 6 = 7770).

Sugerencia: Considere los residuos cuando los siguientes enteros son divididos por *N:* 7, 77, $777, 7777, \ldots, \underbrace{777\cdots77}_{N}$.

4.50 *a)* Demuestre que entre $n + 1$ enteros escogidos al azar, hay dos cuya diferencia es divisible entre n.

b) Demuestre que entre $n + 2$ enteros escogidos al azar, hay dos cuya diferencia es divisible por 2n o hay dos cuya suma es divisible por $2n$.

4.51 Considere una sucesión de N enteros positivos que contiene n enteros distintos, Si $N \geqq 2^n$, demuestre que existe un bloque de enteros consecutivos en la sucesión cuyo producto es un cuadrado perfecto (por ejemplo, en la sucesión 3,7,5,3,7,3,5,7; 7 • 5 • 3 ■ 7 ■ 3 ■ 5 es un cuadrado perfecto).

4.52 Demuestre que de entre $n + 1$ enteros positivos menores o iguales a $2n$ existen dos de ellos que son primos relativos.

4.53 Sean $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots, b_n$ $2n$ números distintos tales que $a_i < b_i$, $i = 1, 2, \ldots, n$. Suponga que $a_1, a_2, \ldots, a_n$ son rearreglados como $a'_1, a'_2, \ldots, a'_n$ tal que $a'_1 > a'_2 > \cdots > a'_n$ y $b_1, b_2, \ldots, b_n$ son rearreglados como $b'_1, b'_2, \ldots, b'_n$ tal que $b'_1 > b'_2 > \cdots > b'_n$. Demuestre que $a'_i < b'_i$, $i = 1, 2, \ldots, n$.

4.54 El siguiente procedimiento puede usarse para seleccionar el n más grande de $2n$ números distintos:

1. Acomode n de los 2« números en orden ascendente. Denótelos por $a_1, a_2, \ldots, a_n$ de manera que $a_1 < a_2 < a_3 < \ldots < a_n$.
2. Acomode los restantes n números en orden descendente. Denótelos por b_1 $b_2, \ldots, b_n$ de manera que $b_1 > b_2 > b_3 > \ldots > b_n$.
3. Compare a_i con b_i y seleccione el mayor de los dos para $i = 1, 2, \ldots, n$.

Por ejemplo, dados los ocho números 4, 2, 7, 6, 5, 3, 1, 8:

1. Acomodamos 4,2, 7, 6 en orden ascendente: 2 < 4 < 6 < 7.
2. Acomodamos 5, 3, 1, 8 en orden descendente: 8 > 5 > 3 > 1.
3. Comparamos 2 y 8; seleccionamos 8. Comparamos 4 y 5; seleccionamos 5. Comparamos 6 y 3; seleccionamos 6. Comparamos 7 y 1; seleccionamos 7.

Demuestre que el procedimiento es correcto.

Sugerencia: Demuestre que exactamente uno de a_i, y b_i está entre el más grande n de los $2n$ números.

CAPÍTULO

CINCO

Grafos y grafos aplanables†

5.1 INTRODUCCIÓN

Como señalamos en el capítulo 4, existen muchos problemas de la vida real que involucran tanto conjuntos discretos de objetos, como relaciones binarias sobre ellos. Por ejemplo, consideremos una serie de encuestas públicas realizadas para determinar la popularidad de los candidatos presidenciales. En cada encuesta se pregunta a los votantes su opinión sobre dos de los candidatos, y se determina un favorito. El resultado de las encuestas se interpreta de la siguiente manera: El candidato *a* le lleva la delantera al candidato *b* si alguna de las siguientes condiciones es verdadera:

1. El candidato *a* va adelante del candidato *b* en una encuesta realizada entre ellos.
2. El candidato *a* va adelante del candidato *c* en una encuesta, y el candidato *c* va adelante del candidato *b* en otra encuesta.
3. El candidato *a* va adelante del candidato *c*, y el candidato *c* va adelante del candidato *d*, y el candidato *d* va adelante del candidato *b* en tres encuestas separadas, y así sucesivamente.

Dados dos candidatos, quisiéramos saber cuándo uno de ellos va adelante de los otros.‡ Sea $S = \{a, b, c, . . .\}$ el conjunto de candidatos y *R* una relación binaria sobre *S* tal que (a, b) está en R si se realiza una encuesta entre *a y b, y a* fue escogido como el candidato favorito. Recordemos que una relación binaria sobre un conjunto puede representarse en forma tabu-

† *N. del E.* Los grafos *aplanables* se conocen también como grafos *planares.*

‡ Observemos que de acuerdo con nuestra interpretación de los resultados de las encuestas, es posible que el candidato *a* lleve la delantera al candidato 6, y que, al mismo tiempo, el candidato *b* también lleve la delantera al candidato *a* (¡vea cómo funciona la política!).

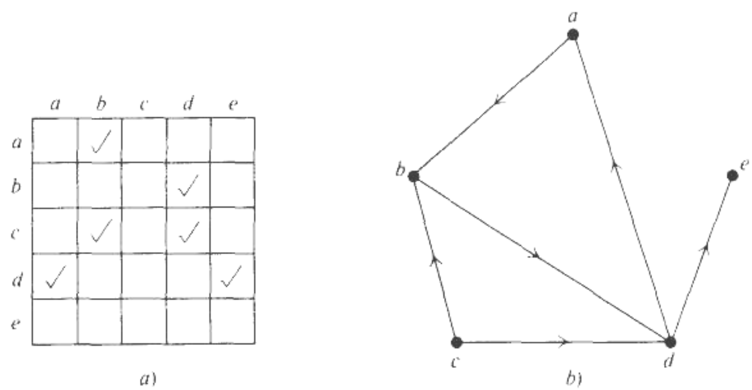


Figura 5.1

lar, como en la figura 5.1 *a*, o en forma gráfica, como en la figura 5.1 *b*. Supongamos que las relaciones binarias en las figuras 5.1*a* y 5.1*b* representan los resultados de las encuestas realizadas. Observamos que el candidato *a* es más popular que el candidato *e* debido a los pares ordenados *(a, b), (b, d), (d, e)* en *R*. Es probable que uno estaría de acuerdo con que la representación gráfica de la relación binaria *R* en la figura 5.1*b* sí es útil al comparar la popularidad de dos candidatos, pues debe existir una "sucesión de flechas" que nos lleven desde el punto correspondiente al candidato más popular hasta el punto correspondiente al candidato menos popular.

Ahora consideremos, como otro ejemplo, un número de ciudades conectadas mediante autopistas. Dado un mapa de las autopistas, podríamos determinar si existe una ruta por autopista entre dos ciudades en el mapa. De igual manera, consideremos todas las listas de posiciones† en una partida de ajedrez. Quisiéramos saber si la lista de posiciones puede alcanzarse a partir de alguna otra lista de posiciones a través de una secuencia de movimientos legales. Es claro que en ambos ejemplos estamos interesados, de nuevo, en objetos discretos y las relaciones binarias sobre ellos. En el ejemplo del mapa de autopistas, sea $S = \{a, b, c, \ldots\}$ el conjunto de ciudades y *R* una relación binaria sobre *S* tal que *(a, b)* está en *R* si existe una autopista de la ciudad a a la ciudad *b*. En el ejemplo de la partida de ajedrez, sea $S = \{a, b, c, \ldots\}$ el conjunto de todas las listas de todas las posiciones y *R* una relación binaria sobre *S* tal que *(a, b)* está en *R* si la lista de posiciones *a* puede transformarse en la lista de posiciones *b* con un solo movimiento legal. Más aún, tanto en estos casos como en el de la popularidad de los candidatos presidenciales, deseamos saber si, dados *a* y *b* en *S* existen $c, d, e, \ldots, h$ en S tales que $\{(a, c), (c, d), (d, e), \ldots, (h, b)\} \subseteq R$.

En efecto, en muchos problemas que incluyen objetos discretos y relaciones binarias, una representación gráfica de los objetos y las relaciones binarias sobre ellos es una forma de representación muy conveniente. Esto nos conduce de modo natural al estudio de la teoría de grafos.

---

† Para ser precisos, una lista de posiciones significa la posición de todas las piezas con una especificación acerca de a quién le corresponde mover (es decir, mueven blancas o negras).

5.2 TERMINOLOGÍA BÁSICA

Un *grafo dirigido G* se define en términos abstractos como un par ordenado (F, *E),* donde *V* es un conjunto y £ es una relación binaria sobre *V.* Como ya se señaló, un grafo dirigido puede representarse geométricamente como un conjunto de puntos marcados *V* con un conjunto de flechas *E* entre parejas de puntos.† Por ejemplo, la figura 5.2 muestra un grafo dirigido. A los elementos de *V* los llamaremos *vértices,* y a los pares ordenados de *E,* aristas del grafo dirigido. Se dice que una arista es *incidente* con los vértices que ella une. Por ejemplo, la arista *(a, b)* es incidente con los vértices *a* y *b.* En ocasiones, cuando deseemos ser más específicos, diremos que la arista *(a, b)* es incidente *desde a* y es incidente *hacia b.* El vértice *a* es llamado el *vértice inicial* y el vértice *b* el *vértice terminal* de la arista *(a, b).* Una arista que es incidente a partir y hacia el mismo vértice, como *(c, c)* en la figura 5.2, es llamado un *lazo.* Se dice que dos vértices son *adyacentes* si están unidos por una arista. Además, al referirnos a una arista *(a, b),* diremos que el vértice *a* es *adyacente* al vértice *b* y también diremos que el vértice *b* es *adyacente desde* el vértice *a.* También que un vértice es *aislado* si no hay una arista incidente con él.

Un *grafo no dirigido G* se define de manera abstracta como un par ordenado *(V, E),* donde *V* es un conjunto y £ es un conjunto de multiconjuntos de dos elementos de *V.* Por ejemplo, *G = {{a, b, c, d}, {{a, b}, {a, d), {b, c), {b, d), {c, c}})* es un grafo no dirigido. Un grafo no dirigido puede representarse geométricamente como un conjunto de puntos marcados *V* con un conjunto de líneas *E* entre los puntos. El grafo no dirigido *G* anterior se muestra en la figura 5.3. Veamos otro ejemplo: sea *V= {a, b, c, d, e}* un conjunto de programas de computadora. La figura 5.4 muestra un grafo no dirigido en el cual hay una arista entre dos vértices si los programas correspondientes comparten algunos datos en común. A partir de este momento, cuando sea claro el contexto, usaremos el término *grafo* para dar a entender que se trata de un grafo dirigido o un grafo no dirigido, o ambos.

Sea *V= {a, b, c, d}* un conjunto de cuatro jugadores en un torneo de tenis de eliminación directa.

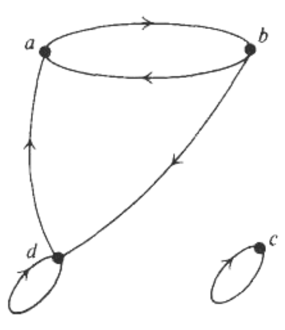


Figura 5.2

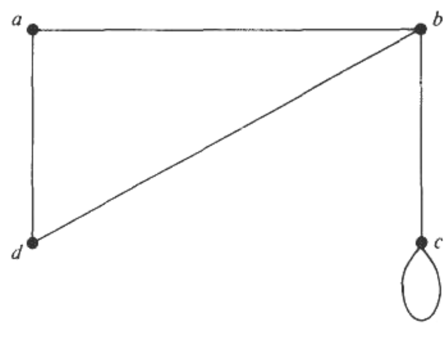


Figura 5.3

† En efecto, podríamos optar para *definir* un grafo dirigido como un conjunto de puntos marcados *V* con un conjunto de flechas entre los puntos, de manera que a lo más existe una flecha desde un punto a otro punto.

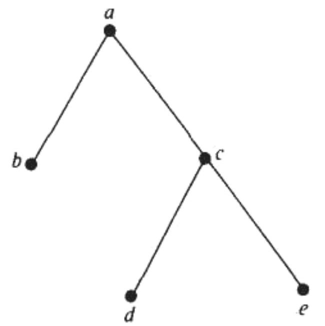


Figura 5.4

Sea $E = \{(a, b), (a, d), (b, d), (c, a), (c, b), (d, c)\}$ una relación binaria sobre V de manera que (x, y) en E significa que x venció a y en el encuentro entre ellos. El grafo $G = (V, E)$ se muestra en la figura 5.5*a*. Sea $V' = \{1, 2, 3, 4\}$ el conjunto de los cuatro capítulos en un libro. Si $E' = \{(1,2), (2,3), (3,1), (3,4), (4,1), (4,2)\}$ es una relación binaria sobre V tal que (1,2) está en E' significa que el material en el capítulo 1 hace referencia al material en el capítulo 2, y así sucesivamente. El grafo $G' = (V', E')$ se muestra en la figura 5.5*b*. Un lector observador habrá reconocido que el grafo de la figura 5.5*b* se parece al grafo de la

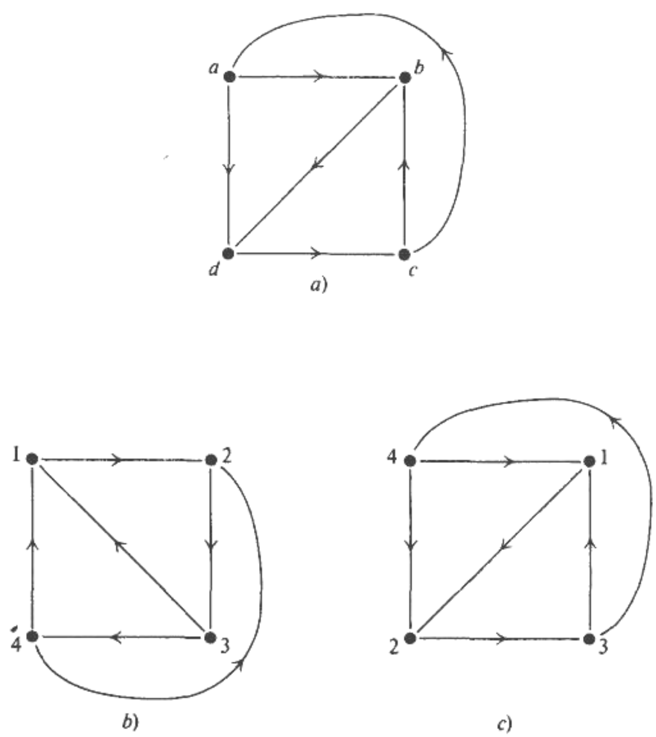


Figura 5.5

figura 5.5*a*. De hecho, la semejanza es más evidente si volvemos a dibujar el grafo de la figura 5.5*b*, como se muestra en la figura 5.5*c*. Si comparamos los grafos de la figura 5.5*a* y figura 5.5*c*, nos damos cuenta que tanto el problema del torneo de tenis como el problema de las referencias cruzadas entre capítulos pueden ser representados de modo abstracto por el mismo grafo. Por tanto, muchos de los resultados referentes a los jugadores de tenis pueden volverse a expresar a partir de los resultados referentes a los capítulos del libro. Por ejemplo, de acuerdo con la figura 5.5*a*, el jugador *b* es mejor que el jugador *d* quien, a su vez, venció al jugador *c*, y de acuerdo con la figura 5.5*c*, el capítulo 1 se refiere al capítulo 2, el cual, a su vez, se refiere al capítulo 3. El concepto de semejanza entre los dos grafos se puede hacer con precisión: dos grafos son *isomorfos* si hay una correspondencia uno a uno entre sus vértices y entre sus aristas, de modo que las incidencias se conservan. En otras palabras, existe una arista entre dos vértices en un grafo si y sólo si hay una arista correspondiente entre los vértices correspondientes en el otro grafo. Por ejemplo, la figura 5.6*a* muestra un par de grafos no dirigidos isomorfos, y la figura 5.6¿ muestra un par de grafos dirigidos isomorfos. En estas dos figuras, los vértices correspondientes en los dos grafos isomorfos están etiquetados con la misma letra, con superíndice de prima y sin superíndice. El lector puede convencerse de que los grafos son isomorfos mediante la comprobación de las relaciones de incidencia.

Sea $G = (V, E)$ un grafo. Diremos que un grafo $G' = \{V', E')$ es un *subgrafo* de G si E' es un subconjunto de E y V' es un subconjunto de V tal que las aristas de E' son incidentes sólo con los vértices de V'. Por ejemplo, la figura 5.7*b* muestra un subgrafo del grafo de la

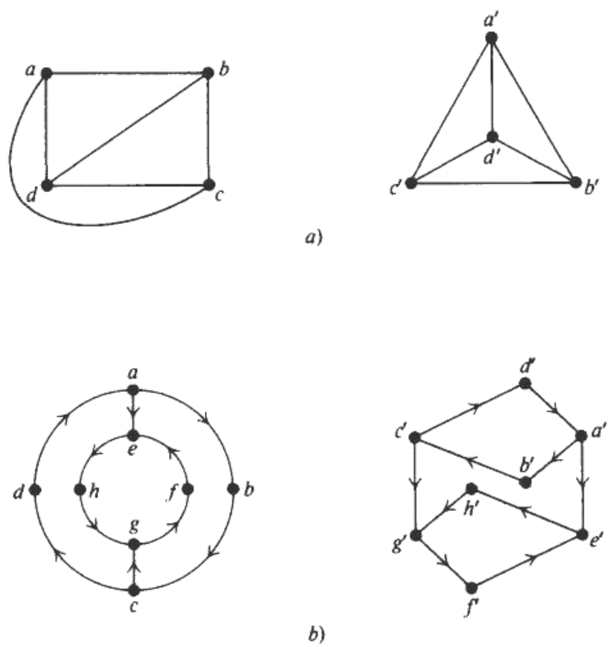


Figura 5.6

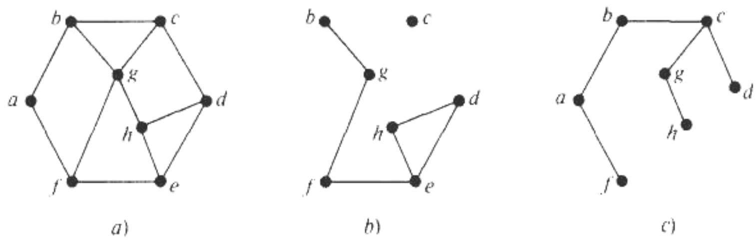


Figura 5.7

figura 5.7*a*. Diremos que un subgrafo de *G* es un *subgrafo generador* si éste contiene todos los vértices de *G*. El *complemento de un subgrafo G' - (V, E*') con respecto al grafo *G* es otro subgrafo *G" = (V", E")* tal que *E"* es igual a *E - E'*, y *V"* sólo contiene a los vértices con los cuales las aristas de *E"* son incidentes.† Por ejemplo, la figura 5.7*c* muestra el complemento del subgrafo de la figura 5.7*b*. El *grafo completo no dirigido* de *n* vértices, denotado por K_n , es un grafo con *n* vértices, en el cual existe una arista entre cada par de vértices distintos. El *complemento de un grafo G* de *n* vértices está definido como su complemento con respecto a K_n y se denota por $\bar{G}$.Por ejemplo, sea *G* un grafo de *n* vértices. Supongamos que los *n* vértices de *G* representan *n* personas, y el conjunto de aristas de *G*, representa una relación de compatibilidad tal que una arista entre dos vértices significa que las dos personas correspondientes pueden trabajar cooperando mutuamente. Es evidente que el conjunto de aristas de $\bar{G}$ representará la relación de incompatibilidad entre las *n* personas. También definimos un grafo *dirigido completo* de *n* vértices como un grafo con *n* vértices en el cual existe exactamente una flecha entre cada par de vértices distintos.

5.3 MULTIGRAFOS Y GRAFOS PESADOS

La definición de un grafo puede extenderse de varias maneras. Sea *G = (V, E)*, donde Fes un conjunto y £ es un multiconjunto de pares ordenados de *V* x *V*. *G* es llamado un *multigrafo dirigido*. Geométricamente, un multigrafo dirigido puede representarse como un conjunto de puntos marcados *V* con un conjunto de flechas *E* entre los puntos, donde no existe restricción en el número de flechas de un punto a otro punto (en efecto, la multiplicidad de un par ordenado de vértices en el multiconjunto *E* es el número de flechas entre los puntos marcados correspondientes). Por ejemplo, la figura 5.8 muestra un multigrafo. Ahora consideremos una representación gráfica de un mapa de carreteras en la cual una arista entre dos ciudades corresponde a un carril en una autopista entre las ciudades. Como a menudo

† De acuerdo con nuestra definición, algunos vértices aislados en *G* que no están incluidos en *G´* tampoco estarán incluidos en *G "*. No obstante, siempre se puede modificar la definición si se desea incluir dichos vértices aislados en *G "*.

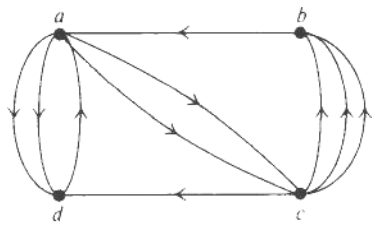


Figura 5.8

hay autopistas de varios carriles entre pares de ciudades, esta representación origina un multigrafo. La noción de un *multigrafo no dirigido* puede definirse de manera similar. De ahora en adelante, cuando sea claro el contexto, usaremos el término *grafo* para significar ya sea un "grafo" o un "multigrafo", o ambos. Por otro lado, cuando sea necesario enfatizar que nos referimos a un grafo (en lugar de a un multigrafo), usaremos el término *grafo lineal.*

Cuando modelamos una situación física con un grafo abstracto, hay muchas ocasiones en las cuales deseamos asociar información adicional a los vértices, a las aristas del grafo o a ambos. Por ejemplo, en un grafo que representa la conexión por autopista entre dos ciudades, podríamos desear asignar un número a cada lado para indicar la distancia entre las dos ciudades conectadas por la arista. También podríamos desear asignar un número a cada vértice para indicar la población de la ciudad. En un grafo que represente los resultados de los encuentros de un torneo de tenis podríamos etiquetar cada arista con las puntuaciones y la fecha del encuentro entre los equipos conectados por la arista. De una manera formal y general, definimos un *grafo pesado* como una cuadrupla ordenada *(V, E,f g),* o una tripleta ordenada *(V, E,f),* o una tripleta ordenada *(V, E, g),* donde *V* es el conjunto de vértices, *E* es el conjunto de aristas,/es una función cuyo dominio es *V, y g* es una función cuyo dominio es *E.* La función/es una asignación de pesos a los vértices, y la función g es una asignación de pesos a las aristas. Los pesos pueden ser números, símbolos, o cualquier cantidad que deseemos asignar a los vértices y a las aristas.

A continuación presentamos algunos ejemplos:

Ejemplo 5.1

Consideremos el problema de modelar el comportamiento de una máquina de ventas, la cual vende caramelos por 15¢ la pieza. Por simplicidad, supongamos que la máquina acepta sólo monedas de cinco y diez centavos y no regresa cambio cuando se han depositado más de 15¢. El grafo pesado de la figura 5.9 es una descripción del comportamiento de la máquina, donde los vértices corresponden a las cantidades que ya han sido depositadas para la venta en un momento dado, a saber, 0, 5, 10 y 15¢ o más. En cualquier momento, un cliente puede hacer una de tres cosas: depositar una moneda de 50, depositar una moneda de 10¢ y presionar un botón para un caramelo de su gusto. En correspondencia, en el grafo de la figura 5.9, hay tres aristas que salen de cada vértice, etiquetados con 5, 10 y *P.* Una arista con peso 5 actualiza la cantidad total depositada en la máquina cuando el cliente introduce una moneda de 5¢, y una arista con peso 10 actualiza la cantidad total depositada en la máquina cuando el cliente introduce una moneda de 10¢. Es claro que cuando estamos en los vértices *a,b* y *c* nada pasará cuando presionemos el botón para seleccionar un caramelo; la máquina liberará un caramelo sólo cuando el vértice *d* ha sido alcanzado.

❒

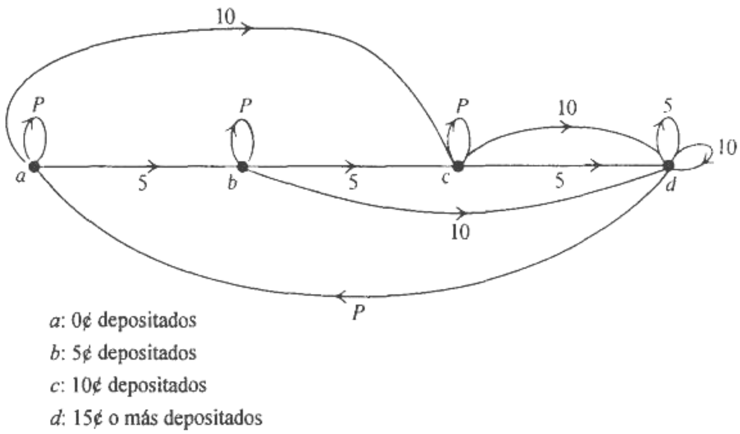


Figura 5.9

Ejemplo 5.2

Consideremos el problema de reconocer enunciados que en lengua inglesa constan de un artículo, seguido de uno, dos o tres adjetivos, un sujeto y al último un verbo, como se muestra enseguida.

El tren se detiene.
Una niña pequeña ríe.
Las nubes grandes, blancas, aborregadas, aparecen.

Cuando examinamos un enunciado palabra por palabra podemos determinar si tiene esta forma especial mediante el seguimiento del grafo pesado de la figura 5.10, comenzando en el vértice *a*. Si el vértice *g* es alcanzado, el enunciado está en la forma especial. Para simplificar el dibujo del grafo, usamos flechas punteadas para indicar el descubrimiento de palabras que se encuentran fuera del orden normal. En tal caso alcanzamos el vértice *h*, que significa la detección de un enunciado "ilegal".

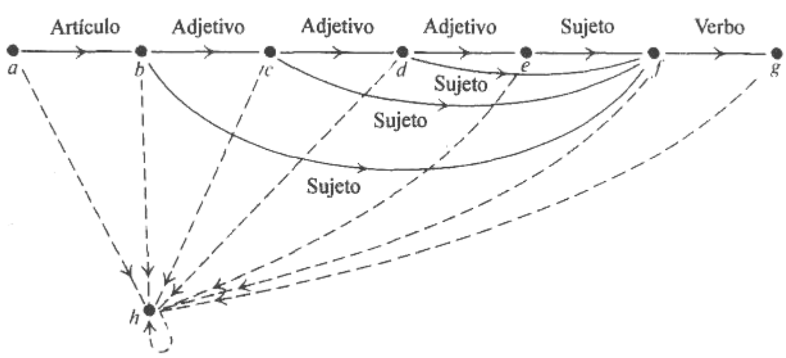


Figura 5.10

Los ejemplos 5.1 y 5.2 ilustran un modelo general que es útil en la descripción de una gran diversidad de problemas físicos, que van desde máquinas automáticas y *hardware* electrónico para computadoras digitales hasta estructuras gramaticales para lenguajes. Dicho modelo es conocido como un *modelo de estado finito,* refiriéndose al hecho de que existe un número finito de vértices en los grafos que describen los problemas. Este tópico lo estudiaremos en el capítulo 7.

5.4 PASEOS Y CIRCUITOS

En un grafo dirigido, *un paseo* es una sucesión de aristas $(e_{i_1}, e_{i_2}, \ldots, e_{i_k})$† tal que el vértice terminal de e_{i_j} coincide con el vértice inicial de $e_{i_{j+1}}$ para $1 \leq j \leq k-1$. Diremos que un paseo es *simple* si no incluye la misma arista dos veces, y que un paseo es *elemental* si no encuentra el mismo vértice dos veces.‡ En la figura 5.11 (e_1, e_2, e_3, e_4) es un paseo; $(e_1, e_2, e_3, e_5, e_8, e_3, e_4)$ es un paseo, pero no es simple; $(e_1, e_2, e_3, e_5, e_9, e_{10}, e_{11}, e_4)$ es un paseo simple, pero no es un paseo elemental. En el ejemplo sobre la popularidad de los candidatos presidenciales de la sección 5.1, que el candidato *a* sea más popular que el candidato *b* significa que existe un paseo del vértice *a* hasta el vértice *b* en el grafo, el cual representa los resultados de las encuestas. En el ejemplo del mapa de autopistas de la sección 5.1, un paseo de un vértice hasta otro vértice en un grafo que representa las conexiones de las autopistas es exactamente una ruta de autopistas entre las ciudades correspondientes.

Un *circuito* es un paseo $(e_{i_1}, e_{i_2}, \ldots, e_{i_k})$ en el cual el vértice terminal e_{i_k} coincide con el vértice inicial e_{i_1}. Diremos que un circuito es *simple* si no incluye la misma arista dos veces, y que un circuíto es *elemental* si no encuentra el mismo vértice dos veces. En la figura 5.11 $(e_1, e_2, e_3, e_5, e_9, e_{10}, e_{12}, e_6, e_7)$ es un circuito simple, pero no es elemental; $(e_1, e_2, e_3, e_5, e_6, e_7)$ es un circuito elemental.

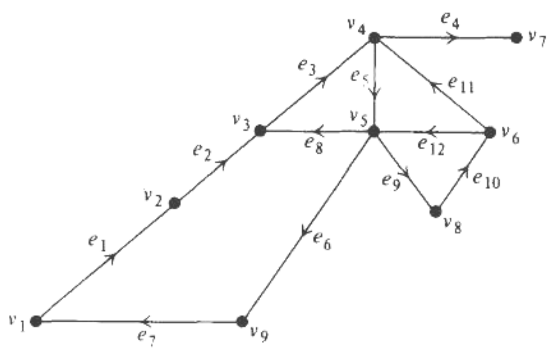


Figura 5.11

† Para simplificar la notación, identificamos las aristas de un grafo mediante nombres de letras como $e_1, e_2, \ldots,$ como se muestra en la figura 5.11.

‡ En otras palabras, no hay dos aristas en la sucesión que tengan el mismo vértice terminal.

En muchas ocasiones también representaremos un paseo o un circuito por la sucesión de vértices que el paseo o circuito encuentra, cuando éste es trazado. Por ejemplo, el paseo (e_1, e_2, e_3, e_4) en el grafo de la figura 5.11 también puede representarse como $(v_1, v_2, v_3, v_4, v_7)$, y el circuito $(e_5, e_9, e_{10}, e_{11})$ puede representarse como $(v_4, v_5, v_8, v_6, v_4)$.

Las nociones de paseos y circuitos en un grafo no dirigido pueden definirse en forma similar. Dejaremos los detalles al lector.

Como se ilustró en ejemplos de las secciones 5.1 y 5.3, existen muchos problemas en los cuales queremos determinar si existe un paseo de un vértice a otro. Ahora presentamos un resultado que es útil para responder la pregunta sobre la existencia de dichos paseos. En la siguiente sección estudiaremos un procedimiento general para encontrar estos paseos.

Teorema 5.1

En un grafo (dirigido o no dirigido) con n vértices, si existe un paseo desde el vértice v_1 hasta el vértice v_2, entonces existe un paseo con no más de $n - 1$ aristas desde el vértice v_1 hasta el vértice v_2.

DEMOSTRACIÓN Supongamos que existe un paseo desde v_1 hasta v_2. Sea $(v_1, \ldots, v_i, \ldots, v_2)$ la sucesión de vértices que el paseo encuentra cuando es trazado desde v_1 hasta v_2. Si existen l aristas en el paseo, entonces existen $l + 1$ vértices en la sucesión. Para l mayor que $n - 1$, debe haber un vértice v_k que aparece más de una vez en la sucesión, es decir, $(v_1, \ldots, v_i, \ldots, v_k, \ldots, v_k, \ldots, v_2)$. Al borrar las aristas en el paseo que conducen a v_k regresando a v_k, tenemos un paseo de v_1 a v_2 que tiene menos aristas que el paseo original. Este argumento puede repetirse hasta que tengamos un paseo con $n - 1$ o menos aristas.

❐

Se dice que un grafo no dirigido es *conexo* si existe un paseo entre cualesquiera dos vértices, y es *no conexo* en otro caso. Se dice que un grafo dirigido es conexo si el grafo no dirigido derivado de éste, al ignorar las direcciones de las aristas, es conexo y es no conexo en otro caso. Entonces se tiene que un grafo no conexo consiste en dos o más componentes, cada una de las cuales es un grafo conexo. Un grafo dirigido es *fuertemente conexo* si para cualesquiera dos vértices a y b en el grafo existe un paseo desde a hasta b, así como también un paseo desde b hasta a. Por ejemplo, la figura 5.12a muestra un grafo conexo que sin embargo no es fuertemente conexo, en tanto que la figura 5.12b muestra un grafo no conexo.

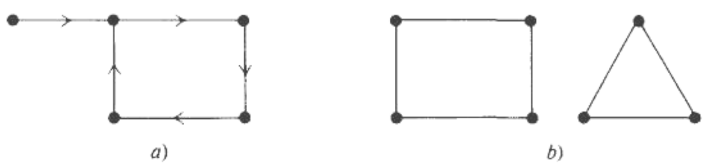


Figura 5.12

5.5 PASEOS CORTOS EN GRAFOS PESADOS

Sea $G = (V, E, w)$ un grafo pesado, donde w es una función de E al conjunto de números reales positivos. Consideremos a V como un conjunto de ciudades y a E como un conjunto de autopistas que conectan estas ciudades. El peso de una arista $\{i,j\}$, se escribe $w(i,j)$† que denota la longitud de la arista $\{i,j\}$, la cual es interpretada como la distancia entre las dos ciudades i y j, a pesar de haber otras interpretaciones como el costo anual t del mantenimiento de la autopista, o el número de accidentes mensuales sobre la autopista, que también son significativas. La *longitud* de un paseo en G se define como la suma de las longitudes de las aristas del paseo. Un problema de gran interés es determinar el paseo más corto de un vértice hasta otro vértice en V. Existen varios procedimientos bien conocidos para solucionar este problema. Aquí presentamos uno descubierto por E. W. Dijkstra [6]. Nuestra presentación es en términos de grafos no dirigidos, no obstante el procedimiento también funciona para grafos dirigidos.

Supongamos que estamos por determinar un paseo corto del vértice a hasta el vértice z en G. En nuestro procedimiento determinamos un paseo más corto desde a hasta algún otro vértice, y luego, un paseo más corto desde a hacia algún otro vértice, y así sucesivamente. Llega el momento en que nuestro procedimiento finaliza, cuando un paseo más corto desde a hasta z es determinado.

Nuestro procedimiento está sustentado en las siguientes observaciones: sea T un subconjunto de vértices de V con $a \notin T$. Sea P el subconjunto de vértices $V - T$. Un paseo más corto de a hasta uno de los vértices de T puede determinarse como sigue: para cada vértice t de T, sea $l(t)$ la longitud del paseo más corto de entre todos los paseos de a hasta t *que no incluyen ningún otro vértice de* T‡ [Observemos que $l(t)$ no necesariamente es la distancia más corta de a hasta t, ya que podría haber un paseo más corto de a hasta t que incluya otros vértices de T.] Llamamos a $l(t)$ el *índice de t con respecto a P*. Entre todos los vértices de T, sea t_x el vértice que posee el índice más pequeño. Aseguramos que la distancia más corta entre a y t_x es, en efecto, igual a $l(t_1)$. Para demostrar nuestra afirmación, supongamos que existe un paseo de a hasta t_l cuya longitud es menor que $l(t_1)$. En ese caso tal paseo debe incluir uno o más de los vértices en $T- \{ t_1\}$. Sea t_2 el primer vértice en $T- \{t_1\}$ que encontramos al trazar este paseo de a hasta t_1. Se sigue que $l(t_2)$ es menor que $l(t_1)$, lo cual es una contradicción. En consecuencia, si al calcular $l(t)$, grabamos la sucesión de vértices del paseo que origina $l(t)$ para cada t en T, también podríamos haber determinado un paseo más corto desde a hasta t_1. Por ejemplo, para el grafo de la figura 5.13*a*, sea $T = \{c, d, e, z\}$. Invitamos al lector a verificar que $l(c) = 3$, $l(d) = 8$, $l(e) = 6$, $l(z) = \infty$. Se sigue que la distancia más corta desde a hasta c es 3.

Sin embargo, también debemos encontrar una manera eficiente de calcular $l\{t)$ para cada t en T. De nuevo, sea T un subconjunto de V y sea $P = V- T$. Suponemos que para cualquier vértice p en P, existe un paseo más corto desde a hasta p que incluye sólo los vértices en P. Suponemos que para cada vértice t en T ya hemos calculado su índice con respecto a P, $l(t)$.

---

† Si queremos ser estrictos, deberíamos usar la notación w({i,y}). No obstante, la notación -$w(i,j)$ es más simple y no es ambigua.

‡ Asignamos $l(t)$ a ∞ si no existe paseo alguno.

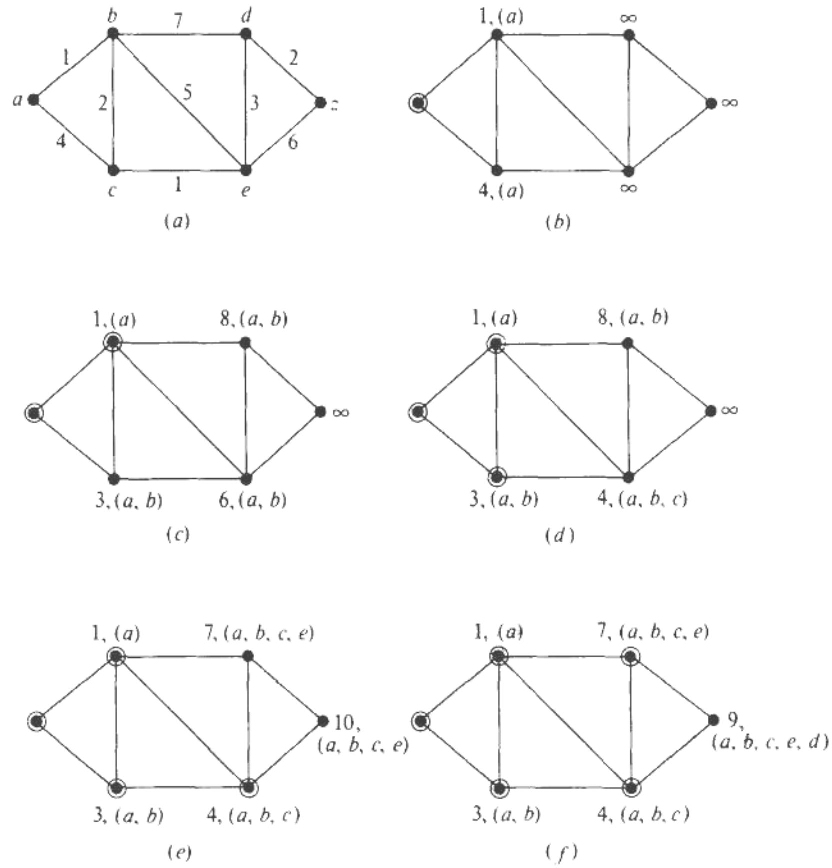


Figura 5.13

Sea x un vértice en T. Sean$P'\ P \cup \{x\}$ y $T'\ T - \{x\}$,y sea $l'(t)$ el índice de un vértice t en T' con respecto a P'. Afirmamos que

$$l'(t) = \text{mín}\ [l(t),\ l(x) + w(x, y)]^{\dagger} \tag{5.1}$$

Para demostrarlo, observamos que existen dos maneras posibles de obtener un paseo más corto desde a hasta t que no incluye ningún vértice en T'. La primera, es tener un paseo que no incluya ni un vértice de T ni tampoco el vértice x. En ese caso, el índice de t con respecto de P' es $l(i)$. La segunda manera es tener un paseo que vaya desde a hasta x que no incluya algún vértice en T, seguido por el lado $\{x, t\}$. En ese caso el índice de t con repecto de P' es $l(x) + w(x, t)$. Debemos señalar que no necesitamos considerar la posibilidad de tener un paseo que va desde a hasta x, luego hacia algún p_1 en P, y entonces hacia t. Si se presentara este caso, mientras exista un paseo más corto desde a hasta p_1 que incluya a x, también existe

† Asignamos $w(x, i)$ a infinito si no existe una arista que una a x y t en el grafo.

uno que no incluye a x y que puede remplazarlo. Por tanto, esto se reduce a la primera posibilidad considerada. Por ejemplo, para el grafo en la figura 5.13*a*, sean $P = \{a, b\}$ y $T = \{c, d, e, z\}$. Supongamos que ya hemos calculado $l(c) = 3$, $l(d) = 8$, $l(e) = 6$, $l(z) = \infty$. Si $P' = \{a, b, c\}$, $T' = \{d, e, z\}$, tenemos

$$l'(d) = \text{mín}\,(8, 3 + \infty) = 8$$
$$l'(e) = \text{mín}\,(6, 3 + 1) = 4$$
$$l'(z) = \text{mín}\,(\infty, 3 + \infty) = \infty$$

Estas observaciones nos conducen al siguiente procedimiento para el cálculo de la distancia más corta desde a hacia cualquier vértice en G:

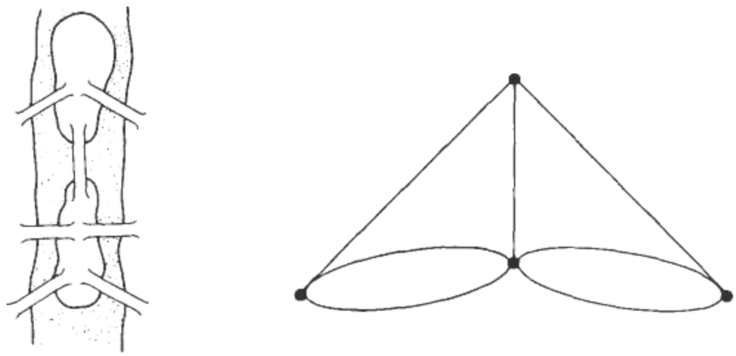
1. En principio, sea $P = \{a\}$ y $T = V - \{a\}$. Para cada vértice t en T, haga $l(t) = w(a, i)$.
2. Seleccione el vértice en T que tiene el índice más pequeño con respecto a P. Denote a este vértice por x
3. Si x es el vértice que deseamos alcanzar desde a, se termina. Si no lo es, haga $P' = P \cup \{x\}$ y $T' = T - \{x\}$. Para cada vértice t en 7", calcule su índice con respecto a P' de acuerdo con (5.1).
4. Repita los pasos 2 y 3, y use P' como P y 7" como T.

Exhortamos al lector a incorporar en este procedimiento los cálculos necesarios para tener un seguimiento del paseo desde a hasta x, cuya longitud es igual a $l(x)$ para cada x en T. En ese caso, estaremos preparados para determinar un paseo más corto desde a hasta x así como la distancia más corta.

Por ejemplo, para el grafo pesado de la figura 5.13*a*, los pasos sucesivos para determinar un paseo más corto entre a y z se muestran desde la figura 5.13 *b*, hasta la figura 5.13/ donde los vértices en P han sido inscritos en una circunferencia. Nos damos cuenta de que los vértices en Testan etiquetados con sus índices. Además, un vértice cuyo índice no es infinito también es etiquetado con la secuencia de vértices correspondientes al paseo desde a hasta el vértice cuya longitud es igual al índice del vértice. Recomendamos al lector esforzarse por comprender completamente cómo estas secuencias de etiquetas de los vértices correspondientes a los paseos desde a son construidas y modificadas. Así, la distancia mínima entre a y z es 9, y también, el paseo más corto es (a, b, c, e, d, z).

5.6 PASEOS EULERIANOS Y CIRCUITOS

Leonhard Euler se convirtió en el padre de la teoría de grafos cuando demostró en 1736 que no era posible cruzar cada uno de los siete puentes sobre el río Pregel en Königsberg, entonces Alemania, hoy Rusia, una y sólo una vez durante una caminata turística. Un mapa de los puentes de Königsberg se muestra en la figura 5.14*a*, el cual puede representarse por el grafo mostrado en la figura 5.14*b*, donde las aristas representan los puentes y los vértices las islas y las dos riberas. Es claro que el problema de cruzar cada uno de los puentes de Königsberg una y sólo una vez es equivalente al de encontrar un paseo en el grafo de la figura 5.146 que pase a través de cada arista una y sólo una vez. Resulta ser que, en lugar de buscar



a) *b)*

Figura 5.14

una solución usando la fuerza bruta por tanteos, lo cual probablemente es lo que los habitantes de Königsberg hicieron, Euler descubrió un criterio muy simple para determinar cuándo existe un paseo en un grafo que pase a través de cada arista una y sólo una vez. Definimos un *paseo euleriano* en un grafo como el paseo que pasa a través de cada lado en el grafo una y sólo una vez. De modo similar, definimos un *circuito euleriano* en un grafo como un circuito que pasa a través de cada arista del grafo una y sólo una vez. Por el momento restringimos nuestro análisis a grafos no dirigidos. Como veremos más adelante, la extensión a grafos dirigidos es prácticamente directa.

Para establecer una condición necesaria y suficiente para la existencia de paseos o circuitos eulerianos en un grafo arbitrario, introducimos la noción del *grado* de un vértice. El grado de un vértice es el número de aristas incidentes en él (señalemos que un lazo contribuiría con 2 al grado de un vértice). Primero observemos que en un grafo cualquiera hay un número par de vértices de grado impar. Puesto que cada arista contribuye con 1 al grado de cada uno de los dos vértices con los cuales es incidente, la suma de los grados de los vértices es igual al doble del número de aristas de un grafo. De esto se sigue que debe existir un número par de vértices de grado impar.

La existencia de paseos eulerianos o circuitos eulerianos en un grafo está relacionada con los grados de los vértices. Ahora mostramos un resultado debido a Euler:

Teorema 5.2 Un grafo no dirigido tiene un paseo euleriano si y sólo si éste es conexo y tiene cero o dos vértices de grado impar.†

DEMOSTRACIÓN Supongamos que el grafo tiene un paseo euleriano. Que el grafo debe ser conexo es obvio. Cuando el paseo euleriano es trazado, observamos que cada vez que el paseo cruza un vértice, pasa a través de dos aristas las cuales son incidentes con el vértice y no han sido recorridas anteriormente. Así, excepto para los dos vértices en los dos extremos del paseo, el grado de cualquier vértice en el grafo debe ser par. Si los dos vértices en los extremos del paseo euleriano son distintos, ellos son los únicos dos

† Descartamos e! caso de los grafos que contienen vértices aislados por ser poco interesante.

vértices con grado impar. Si ellos coinciden, todos los vértices tienen grado par, y el paseo euleriano es un circuito euleriano. Así, la condición de necesidad se ha demostrado.

Para demostrar la condición de suficiencia, construimos un paseo euleriano que inicie en uno de los dos vértices que son de grado impar† y continuamos a través de las aristas del grafo de manera tal que no pasemos más de una vez a través de alguna arista. Para un vértice de grado par, siempre que un paseo "entra" al vértice a través de una arista, siempre podrá "salir" del vértice por otra arista que no ha sido recorrida con anterioridad. Por tanto, cuando la construcción llegue a su fin, deberemos haber alcanzado el otro vértice de grado impar. Si todas las aristas del grafo se recorrieron de esta manera, es obvio que tenemos un paseo euleriano. Si no fueron recorridas todas las aristas del grafo, entonces eliminemos aquellas aristas que han sido recorridas y obtendremos un subgrafo formado por las aristas restantes. Todos los grados de los vértices de este subgrafo son pares. Además, este subgrafo debe tocar al paseo que habíamos recorrido en uno o más vértices, ya que el grafo original es conexo. Al comenzar desde alguno de estos vértices, podemos construir de nuevo un paseo que pase a través de las aristas. Debido a que los grados de todos los vértices son pares, este paseo debe regresar finalmente al vértice en el cual comenzó. Podemos combinar este paseo con el que ya teníamos construido para obtener el paseo que comienza y termina en los dos vértices de grado impar. Si es necesario, el argumento se repite hasta obtener un paseo que pase a través de todas las aristas del grafo. □

Corolario 5.2.1 Un grafo no dirigido tiene un circuito euleriano si y sólo si es conexo y sus vértices son todos de grado par.

Concluimos a partir del teorema 5.2 y el corolario 5.2.1 que el grafo de la figura 5.15*a* tiene un paseo euleriano, pero no tiene un circuito euleriano debido a que el grafo es conexo y

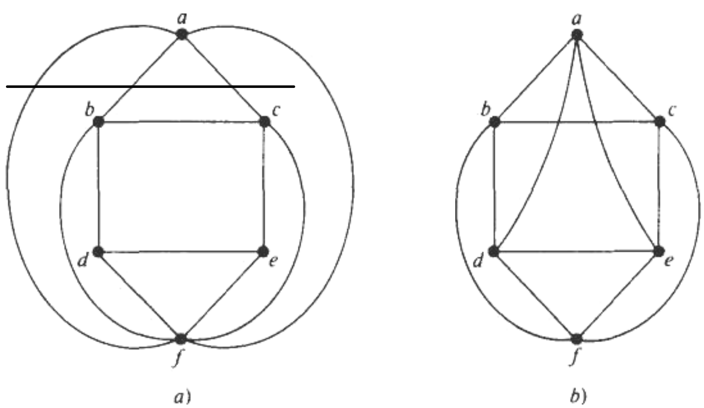


Figura 5.15

† Comenzamos en un vértice arbitrario si no hay un vértice de grado impar.

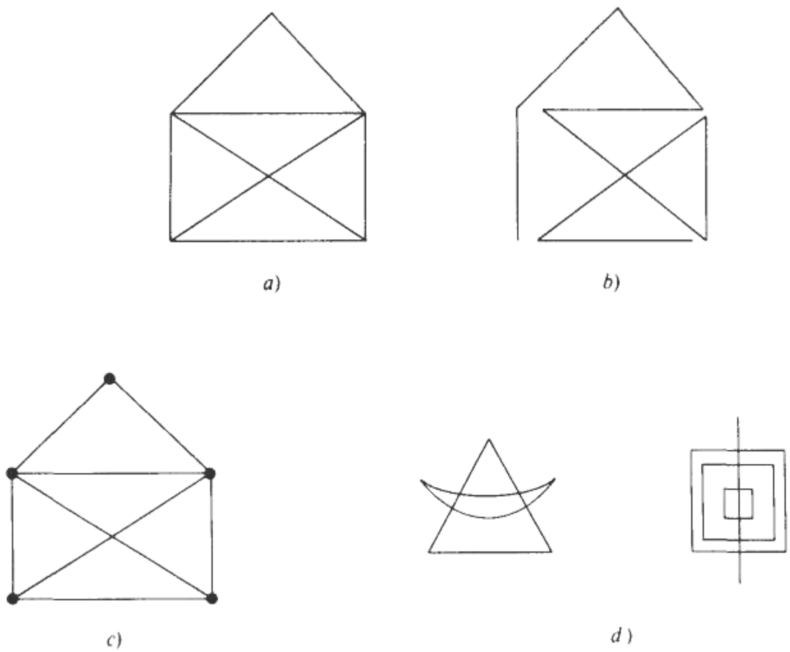


Figura 5.16

tiene exactamente dos vértices, *d* y *e*, de grado impar. También el grafo de la figura 5.156 tiene un circuito euleriano, debido a que el grafo es conexo y todos los vértices son de grado par.

Ahora mostraremos algunos ejemplos ilustrativos.

Ejemplo 5.3

A menudo encontramos el reto de determinar la posibilidad de dibujar una figura dada con un trazo continuo de manera que ninguna parte de la figura sea repetida. Por ej emplo, la figura que se muestra en la figura 5.16*a* puede trazarse como se muestra en la figura 5.16*b*. Imagine que en la figura 5.16*a* se muestra un grafo como el de la figura 5.16*c*. Entonces el problema de trazar la figura 5.16*a* es determinar la existencia de un paseo euleriano en el grafo de la figura 5.16*c*. Puesto que el grafo de la figura 5.16*c* tiene sólo dos vértices de grado impar, entonces tiene un paseo euleriano. De modo similar, las dos figuras de la figura 5.16*d* pueden dibujarse por un trazo continuo sin que se repita ninguna parte de las figuras.

❐

Ejemplo 5.4

Deseamos saber si es posible arreglar las 28 fichas diferentes de un dominó† en un círculo de manera que las mitades adyacentes de cualesquiera dos fichas adyacentes en el arreglo sean las mismas. Construimos un grafo con siete vértices correspondientes a blanco, 1, 2, 3, 4, 5 y 6. Existe una arista entre cualesquier par de vértices correspon-

† Véase el ejemplo 3.17.

dientes a una ficha cuyas dos mitades son los dos vértices con los que la arista es incidente. Se tiene que un circuito euleriano en este grafo corresponderá a un arreglo circular, como se especificó anteriormente. Dado que el grado de cualquier vértice en este grafo es 8, existe, en efecto, un circuito euleriano.

Nuestros resultados pueden extenderse de inmediato para grafos dirigidos. En un grafo dirigido, el *grado de entrada* de un vértice es el número de aristas que son incidentes hacia éste, y el *grado de salida* de un vértice es el número de aristas que son incidentes desde éste. De modo similar a los resultados sobre grafos no dirigidos tenemos lo siguiente:

Teorema 5.3

Un grafo dirigido tiene un circuito euleriano si y sólo si es conexo y el grado de entrada de cualquier vértice es igual a su grado de salida. Un grafo dirigido tiene un paseo euleriano si y sólo si es conexo y el grado de entrada de cualquier vértice es igual a su grado de salida, con la posible excepción de dos vértices. Para estos dos vértices, el grado de entrada de uno de ellos es mayor en uno que su grado de salida, y el grado de entrada del otro es menor en uno que su grado de salida.

Ejemplo 5.5

Un ejemplo interesante proviene de un problema de conversión analógica a digital. La superficie de un tambor giratorio se divide en 16 sectores, como se muestra en la figura 5.17*a*. La información posicional del tambor se representa mediante las señales digitales binarias, *a,b,c d*, como se muestra en la figura 5.17*b*, donde materiales conductores (área sombreada) y no conductores (área en blanco), son usados para construir los sectores. De acuerdo con la posición del tambor, las terminales *a,b,c* y *d* se encontrarán conectadas a tierra o bien aisladas de ésta. Por ejemplo, cuando la posición del tambor es la mostrada en la figura 5.17*b*, las terminales *a* , *c* y *d* están conectadas a tierra, en tanto que la terminal *b* no lo está. Para que las 16 posiciones diferentes del tambor sean representadas diferentemente por las señales binarias de las terminales, los sectores deben construirse de tal manera que dos patrones de conductores y no conductores de

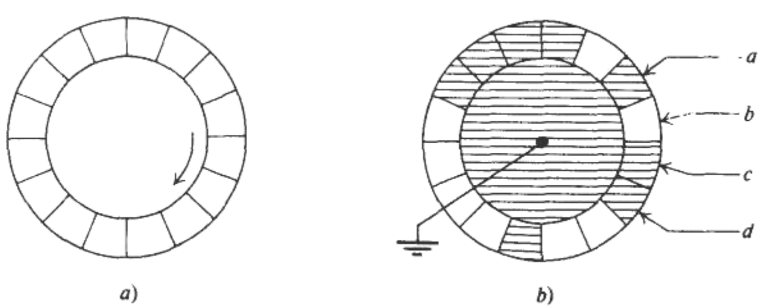


Figura 5.17

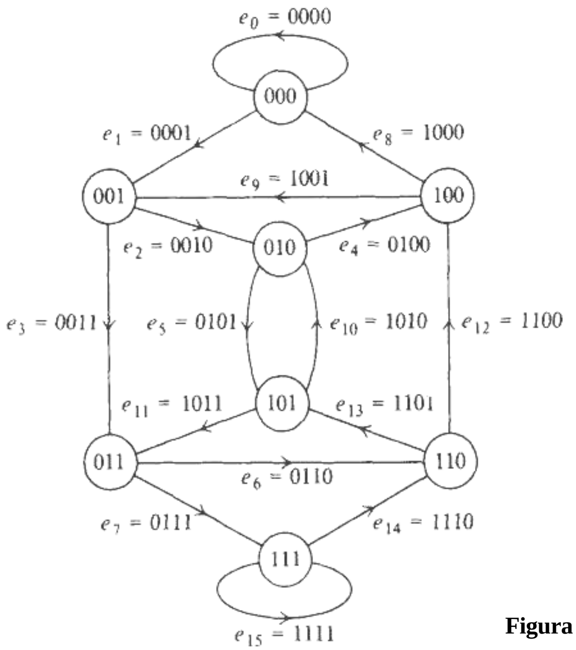


Figura 5.18

cuatro sectores consecutivos no sean los mismos. El problema es determinar si tal arreglo de sectores conductores y no conductores existe y, si es así, determinar de qué arreglo se trata. Si asignamos el dígito binario 0 para denotar un sector conductor y el dígito binario 1 para denotar un sector no conductor, podemos replantear el problema de la siguiente manera: acomodar 16 dígitos binarios en un arreglo circular tal que las 16 sucesiones de 4 dígitos consecutivos sean todas distintas.

La respuesta a la pregunta de la posibilidad de tal arreglo es afirmativa, y esto es realmente obvio una vez que se toma el punto de vista adecuado. Construyamos un grafo dirigido con ocho vértices, los cuales son etiquetados con los ocho diferentes números binarios de 3 dígitos {000, 001,. . ., 111}. A partir de un vértice etiquetado $\alpha_1\alpha_2\alpha_3$, existe una arista hacia el vértice etiquetado $\alpha_2 \alpha_3 0$ y una arista hacia el vértice etiquetado $\alpha_2 \alpha_3$. El grafo así construido se muestra en la figura 5.18. Además, etiquetemos cada arista del grafo con un número binario de 4 dígitos. En particular, la arista del vértice $\alpha_1\alpha_2\alpha_3$ al vértice $\alpha_2\alpha_3 0$ es etiquetado como $\alpha_1\alpha_2\alpha_3 0$, y la arista del vértice $\alpha_1 \alpha_2 \alpha_3$ al vértice $\alpha_2\alpha_3 1$ es etiquetada como $\alpha_1\alpha_2\alpha_3 1$., Puesto que los vértices son etiquetados con los ocho números binarios distintos de 3 dígitos, las aristas serán etiquetadas con los 16 números binarios distintos de 4 dígitos. En un paseo del grafo, las etiquetas para cualesquiera dos aristas consecutivas deben ser de la forma $\alpha_1\alpha_2\alpha_3\alpha_4$ y $\alpha_2\alpha_3\alpha_4\alpha_5$; es decir, los tres dígitos finales de la etiqueta de la primera arista son idénticos a los tres dígitos iniciales de la etiqueta de la segunda arista. Puesto que los 16 lados en el grafo son etiquetados con distintos números binarios, se tiene que existe un arreglo circular de 16 dígitos binarios, que corresponde al circuito euleriano del grafo, en el cual todas las sucesiones de 4 dígitos consecutivos son distintas. Por ejemplo, para el circuito euleriano $(e_0, e_1, e_2, e_5, e_{10}, e_4, e_9, e_3, e_6, e_{13}, e_{11}, e_7, e_{15}, e_{14}, e_{12}, e_8)$, la sucesión correspondiente de 16 dígitos binarios es 0000101001101111 (el arreglo circular se obtiene al cerrar los dos extremos de la sucesión). De acuerdo con nuestros resultados hasta ahora obtenidos, la existencia de un circuito euleriano en el grafo es obvia, debido a que cualquiera de los vértices tiene grado de entrada igual a 2 y un grado de salida igual a 2. Además, podemos encontrar un circuito euleriano en el grafo, si seguimos el procedimiento de construcción sugerido en la demostración del teorema 5.2.

Usando un argumento similar, podemos demostrar que es posible arreglar *2''* dígitos binarios en un arreglo circular tal que las 2^n sucesiones de *n* dígitos consecutivos en el arreglo sean todas distintas. Para demostrar esto, construimos un grafo dirigido con 2«-i vértices, los cuales son etiquetados con los 2^{n-1} números binarios de (n - 1) dígitos. A partir del vértice $\alpha_1\alpha_2\alpha_3 \cdots \alpha_{n-1}$, existe una arista al vértice $\alpha_2\alpha_3 \cdots \alpha_{n-1}0$, el cual es etiquetado como $\alpha_1\alpha_2\alpha_3 \cdots \alpha_{n-1}0$, y una arista hacia el vértice $\alpha_2\alpha_3 \cdots \alpha_{n-1}1$, el cual es etiquetado como $\alpha_1\alpha_2\alpha_3 \cdots \alpha_{n-1}1$.. Es claro que este grafo tiene un circuito euleriano que corresponde a un arreglo circular de los 2^n dígitos binarios.

□

5.7 PASEOS Y CIRCUITOS HAMILTONIANOS

Un problema similar a la determinación de un paseo o circuito euleriano, es el de determinar un paseo o un circuito que pasa a través de cada vértice en un grafo una y sólo una vez. Definimos un *paseo* (circuito) *hamiltoniano* como un paseo (circuito) que pasa a través de cada uno de los vértices de un grafo exactamente una vez (Sir William Hamilton inventó el juego "alrededor de todo el mundo" en el cual el jugador es invitado a determinar una ruta a lo largo de un dodecaedro tal que pasará a través de cada vértice una y sólo una vez).

Como un ejemplo, consideremos el problema de sentar un grupo de personas alrededor de una mesa. Si consideramos que los vértices de un grafo no dirigido denotan las personas y las aristas denotan cuando dos personas sean amigas, un circuito hamiltoniano corresponde a la manera de sentarlas de modo que cada una tenga un amigo a cada lado.

Aunque el problema de determinar la existencia de circuitos o paseos hamiltonianos tiene la misma connotación que el de determinar la existencia de paseos o circuitos eulerianos, no se conoce ninguna condición necesaria y suficiente. Para demostrar que un grafo dado tiene un paseo o circuito hamiltoniano, podemos apoyarnos en una construcción explícita de dicho paseo o circuito. Por ejemplo, en el grafo de la figura 5.19 las aristas gruesas constituyen un circuito hamiltoniano.

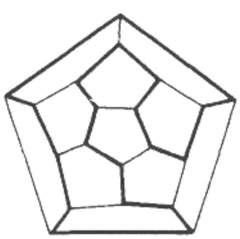


Figura 5.19

Mostraremos ahora algunos resultados generales sobre la existencia de paseos o circuitos hamiltonianos. Nuestro primer resultado es más bien una condición general que es *suficiente* para garantizar la existencia de un paseo hamiltoniano en un grafo no dirigido:

Teorema 5.4 Sea G un grafo lineal de n vértices. Si la suma de los grados para cada par de vértices de G es n - 1 o mayor, entonces existe un paseo hamiltoniano en G}

DEMOSTRACIÓN Primero demostraremos que G es un grafo conexo. Supongamos que G tiene dos o más componentes no conexas. Sea v, un vértice en una componente que tiene n_1 vértices y v_2 un vértice en otra componente que tiene n_2 vértices. Debido a que el grado de v_1 es a lo más $n_1 - 1$ y el grado de v_2 es a lo más $n_2 - 1$, la suma de sus grados es a lo más $n_1 + n_2 - 2$, que es menor que $n - 1$, y tenemos una contradicción.

Ahora demostraremos cómo puede ser construido paso por paso un paseo hamiltoniano, comenzando con un paseo que contiene una sola arista. Supongamos que existe un paseo de $p - 1$ aristas, $p < n$, en G el cual cruza la sucesión de vértices $(v_1\ v_2, \ldots, v_p)$. Si v_1 o v_p es adyacente a un vértice que no está en el paseo, podemos extender de inmediato el paseo para incluir este vértice y obtener un paseo de p aristas. En otro caso, tanto v_1 como v_p son adyacentes solamente a los vértices que están en el paseo. Queremos demostrar que en ese caso existe un circuito que contiene exactamente los vértices $v_1, v_2, \ldots, v_p$. Si v_1 es adyacente a v_p, entonces el circuito $(v_1, v_2, \ldots, v_p, v_1)$ será suficiente, de manera que consideraremos que v_1 sólo es adyacente a $v_{i_1}, v_{i_2}, \ldots, v_{i_k}$, donde $2 \leqq i_j \leqq p - 1$.‡t Si v_p es adyacente a alguno de $v_{i_1-1}, v_{i_2-1}, \ldots, v_{i_k-1}$, digamos a v_{j-1}, entonces, como se muestra en la figura 5.20, el circuito $(v_1, v_2, v_3, \ldots, v_{j-1}, v_p, v_{p-1}, \ldots, v_j, v_1)$ contiene exactamente los vértices $v_1, v_2, \ldots, v_p$. Si v_p no es adyacente a ninguno de $v_{i_1-1}, v_{i_2-1}, \ldots, v_{i_k-1}$, entonces v_p es adyacente a lo más $ap - k - 1$ vértices. En consecuencia, la suma de los grados de v_1 y v_p es a lo más $n - 2$, lo cual es una contradicción.

Ahora que tenemos un circuito que contiene a todos los vértices $v_1, v_2, \ldots, v_p$, tomemos un vértice v_x, que no esté en el circuito. Debido a que G es conexo, existe un vértice v_k que no está en el circuito con una arista entre v_x y v_k para algún v_k en $\{v_1, v_2,$

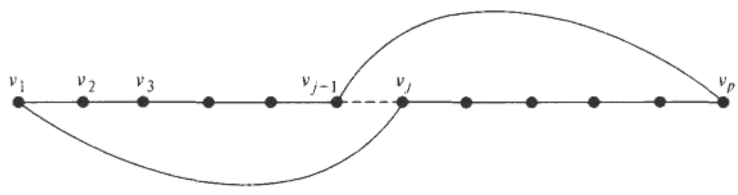


Figura 5.20

† Efectivamente, $i_t = 2$.

‡ Véase el problema 5.33 para una generalización de este resultado.

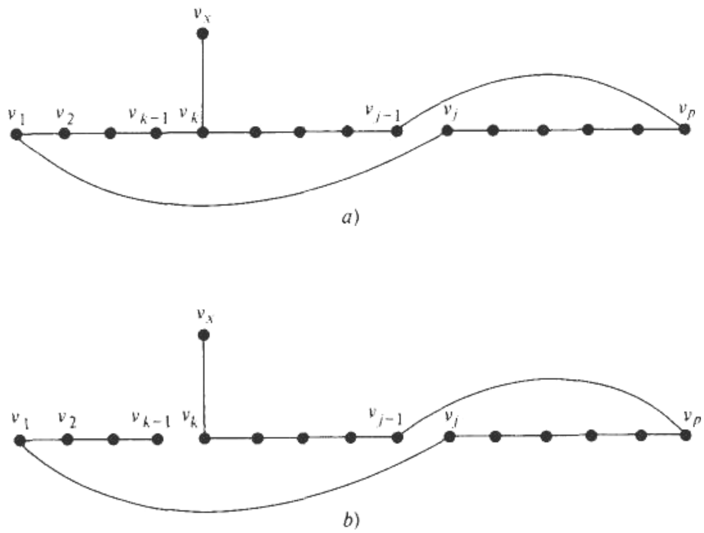


Figura 5.21

$\ldots, v_p\}$, como se muestra en la figura 5.21*a*. Ahora tenemos el paseo $(v_x, v_k, v_{k+1}, \ldots, v_{j-1}, v_p, v_{p-1}, \ldots, v_j, v_1, v_2, \ldots, v_{k-1})$, el cual contiene *p* aristas, como se muestra en la figura 5.216.

Podemos repetir la construcción anterior hasta que obtengamos un paseo con n - 1 aristas. ❐

Es fácil ver que la condición del teorema 5.4 es una condición suficiente pero no necesaria para la existencia de un paseo hamiltoniano en un grafio. Sea *G* un n-ágono, $n > 5$. Es claro que *G* tiene un paseo hamiltoniano a pesar de que la suma de los grados de dos vértices cualesquiera es 4.

Ejemplo 5.6

Consideremos el problema de programar siete exámenes en siete días, de manera que dos exámenes aplicados por el mismo instructor no se programen en días consecutivos. Si ningún instructor aplica más de cuatro exámenes, mostraremos que siempre será posible programar los exámenes. Sea *G* un grafo con siete vértices que corresponden a los siete exámenes. Existe una arista entre cualesquiera dos vértices, la cual corresponde a dos exámenes aplicados por diferentes instructores. Puesto que el grado de cada vértice es al menos 3, la suma de los grados de cualesquiera dos vértices es al menos 6. En consecuencia, G siempre contendrá un paseo hamiltoniano, el cual corresponde a una programación adecuada para los siete exámenes. ❐

Otro resultado interesante es:

Teorema 5.5

Siempre existe un paseo hamiltoniano en un grafo dirigido completo.

DEMOSTRACIÓN Consideremos que existe un paseo con p - 1 aristas en un grafo dirigido completo, el cual craza la sucesión de vértices (v_1 v_2,..., v_p). Sea v_x un vértice que no está incluido en este paseo. Si existe una arista (v_x, v_1) en el grafo, podemos aumentar el paseo original agregando la arista *(v_x, v{)* al paseo, de manera que el vértice v_x estará incluido en el paseo aumentado. Si, por otro lado, no existe una arista de v_x a v_1 entonces debe existir una arista (v_1, v_x) en el grafo. Supongamos que (v_x, v_2) también es una arista del grafo. Podemos remplazar la arista (v_1 v_2) en el paseo original con dos aristas (v_1, v_x) y (v_x, v_2) de manera que el vértice v_x estará incluido en el paseo aumentado. Por otro lado, si no existe una arista de v_x a v_2, entonces debe existir una arista (v_2, v_x) en el paseo y podemos repetir el argumento. Por último, si encontramos que no es posible incluir el vértice v_x en cualquier paseo aumentado mediante el remplazo de una arista (v_k, v_{k+1}) en el paseo original con dos aristas (v_k, v_x) y (v_x, v_{k+1}) con $1 \leqq k \leqq p-1$, entonces concluimos que debe existir una arista (v_p, v_x) en el grafo. Podemos, por tanto, aumentar el paseo original agregándole la arista (v_p, v_x) de manera que el vértice v_x será incluido en el paseo aumentado. Podemos repetir el argumento hasta que todos los vértices del grafo sean incluidos en el paseo. ❐

Ejemplo 5.7

Como una aplicación del resultado del teorema 5.5, consideremos el problema de clasificar a los jugadores de un torneo de tenis de eliminación directa, de modo que el jugador *a* se clasificará mejor que el jugador *b* si *a* vence a *b,* o si *a* vence al jugador que venció a *b,* o si *a* vence a un jugador que venció a otro jugador que venció a su vez a *b,* y así sucesivamente. Como los resultados de los encuentros pueden representarse como un grafo dirigido completo, la existencia de un paseo hamiltoniano en el grafo significa que siempre es posible clasificar a los jugadores linealmente (observe que dicha clasificación no es necesariamente única). ❐

Ejemplo 5.8

Otro ejemplo ilustrativo consiste en considerar el problema de imprimir y luego encuadernar *n* libros. Existe una impresora y una encuadernadora. Consideremos que p_i y b_i denotan el tiempo de impresión y el tiempo de encuademación del libro *i*, respectivamente. Si para dos libros cualesquiera *i, j* sabemos que bien $b_i \geqq p_j$ o bien $b_j \geqq p_i$, demostraremos que es posible especificar el orden en el cual los libros son impresos (y luego encuadernados), de manera que la máquina encuadernadora se mantendrá ocupada hasta que todos los libros sean encuadernados, una vez que el primer libro es impreso (así, el tiempo total que toma completar toda la tarea es $p_k + \Sigma_{i=1}^{n} b_i$ para algún *k).* Construiremos un grafo dirigido de n vértices correspondientes a los *n* libros. Existe una arista del vértice *i* al vértice y si y sólo si $b_i \geqq p_j$. Observamos que éste es un grafo dirigido completo, y un paseo hamiltoniano en el grafo dirigido completo será un ordenamiento de los libros, que satisface la condición anteriormente establecida. ❐

No existe un método general para la solución del problema consistente en demostrar la no existencia de un paseo o circuito hamiltoniano en un grafo. Aquí presentamos, no obstante, un ejemplo ilustrativo:

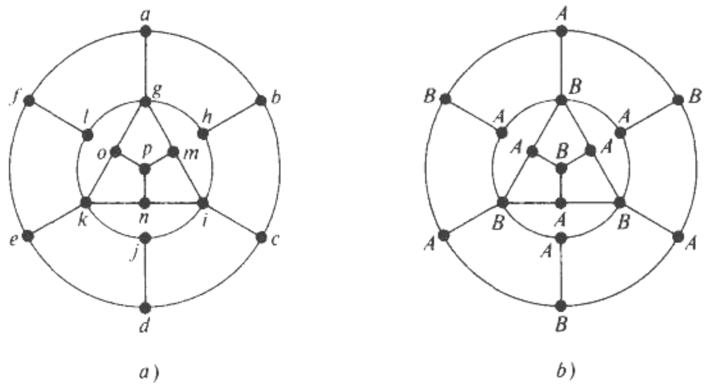


Figura 5.22

Ejemplo 5.9

Queremos demostrar que el grafo de la figura 5.22*a* no tiene un paseo hamiltoniano. Etiquetamos el vértice *a* como *A* y etiquetamos todos los vértices que le son adyacentes como *B.* A continuación etiquetamos todos los vértices adyacentes a un vértice *B* como *A,* y etiquetamos todos los vértices adyacentes a un vértice *A* como *B,* hasta que todos los vértices sean etiquetados. El grafo etiquetado se muestra en la figura 5.22*b.* Si existe un paseo hamiltoniano en el grafo, entonces éste debe pasar a través de los vértices *A y* los vértices *B* alternativamente. No obstante, ya que hay nueve vértices *A y* siete vértices *B,* la existencia del paseo hamiltoniano es imposible. □

* 5.8 EL PROBLEMA DEL AGENTE VIAJERO

En esta sección vamos a analizar una extensión natural del problema de encontrar un circuito hamiltoniano en un grafo. El *problema del agente viajero* que ha sido durante mucho tiempo de gran interés. Sea $G = (V, E, w)$ un grafo completo de *n* vértices, donde *w* es una función de *E* hacia el conjunto de números reales positivos tal que para cualesquiera tres vértices i, j, k en V

$$w(i, j) + w(j, k) \geqq w(i, k)^{\dagger}$$

Nos referimos a $w(i,j)$ como la longitud de la arista $\{i,j\}$. El problema del agente viajero requiere de un circuito hamiltoniano de longitud mínima, donde, nuevamente, la longitud del circuito está definida como la suma de las longitudes de las aristas del circuito.

Una interpretación física de la formulación abstracta es más obvia: consideremos el grafo *G* como un mapa de *n* ciudades donde $w(i,j)$ es la distancia entre las ciudades *i* y *j.* Un

† Esta condición, conocida como la *desigualdad del triángulo,* es necesaria para demostrar el resultado de (5.2). Por otro lado, uno puede tratar de argumentar que la desigualdad del triángulo se satisface en la mayoría de los casos de la vida real.

agente desea realizar un recorrido por las n ciudades, que comience y termine en la misma ciudad, e incluya visitar a cada una de las n -1 ciudades restantes una y sólo una vez. Además, se desea un itinerario que tenga una distancia total mínima. El problema del agente viajero resulta ser difícil en el sentido de que no se conoce un procedimiento "eficiente"† para resolverlo. Al recordar nuestro análisis de la sección 4.7 sobre el problema de la programación de tareas, desearíamos buscar procedimientos sencillos que proporcionaran buenos resultados en el problema del agente viajero. Para ejemplificar esta posibilidad, presentamos un procedimiento conocido como el *método del vecino más cercano*, que proporciona resultados razonablemente buenos para el problema del agente viajero:

1. Comience con un vértice escogido de modo arbitrario, y encuentre el vértice que esté más cercano al vértice inicial para formar un paseo inicial de una arista. Aumentaremos este paseo en un proceso de vértice por vértice como es descrito en el paso 2.
2. Denote como x al último vértice que fue agregado al paseo. De entre todos los vértices que no están en el paseo, seleccione aquel que esté más cercano a x, y agregue al paseo la arista que conecta x con este vértice. Repita este paso hasta que todos los vértices de G estén incluidos en el paseo.
3. Forme un circuito agregando la arista que conecta el vértice inicial y el último vértice agregado.

Por ejemplo, para el grafo mostrado en la figura 5.23*a*, si comenzamos desde el vértice *a*, una construcción vértice por vértice de un circuito hamiltoniano de acuerdo con el método del vecino más cercano se muestra en la figura 5.23*b*, hasta la figura 5.23*e*. Observemos que la distancia total de este circuito es 40, en tanto que la distancia total de un circuito hamiltoniano mínimo es 37, como se observa en la figura 5.23*f*.

Demostraremos el siguiente resultado:

Teorema 5.6

Para un grafo con n vértices, sea d la distancia total de un circuito hamiltoniano obtenido de acuerdo con el método del vecino más cercano y sea d_0 la distancia total de un circuito hamiltoniano mínimo. Entonces,

$$\frac{d}{d_0} \leq \frac{1}{2}\lceil \lg n \rceil + \frac{1}{2} \text{‡} \tag{5.2}$$

DEMOSTRACIÓN Antes de proceder con la demostración, daremos la idea general de la demostración considerando un caso específico. Sea D un circuito hamiltoniano obtenido de acuerdo con el método del vecino más cercano. Sean l_1 la longitud de la

† Como podría responder el lector, esto depende por completo de lo que entendamos por un procedimiento "eficiente". La eficiencia de los procedimientos de computación es un tópico importante en teoría de la computación, la cual estudiaremos en el capítulo 8. Por ahora, digamos solamente que no existe un procedimiento conocido que pueda resolver el problema del agente viajero con más de varios cientos de ciudades mediante el uso de una computadora digital de gran escala en una cantidad de tiempo razonable.

‡ Recordemos que adoptamos la notación lg n para $\log_2 n$. Usamos *[x]* para denotar el entero más pequeño mayor que o igual *ax*.

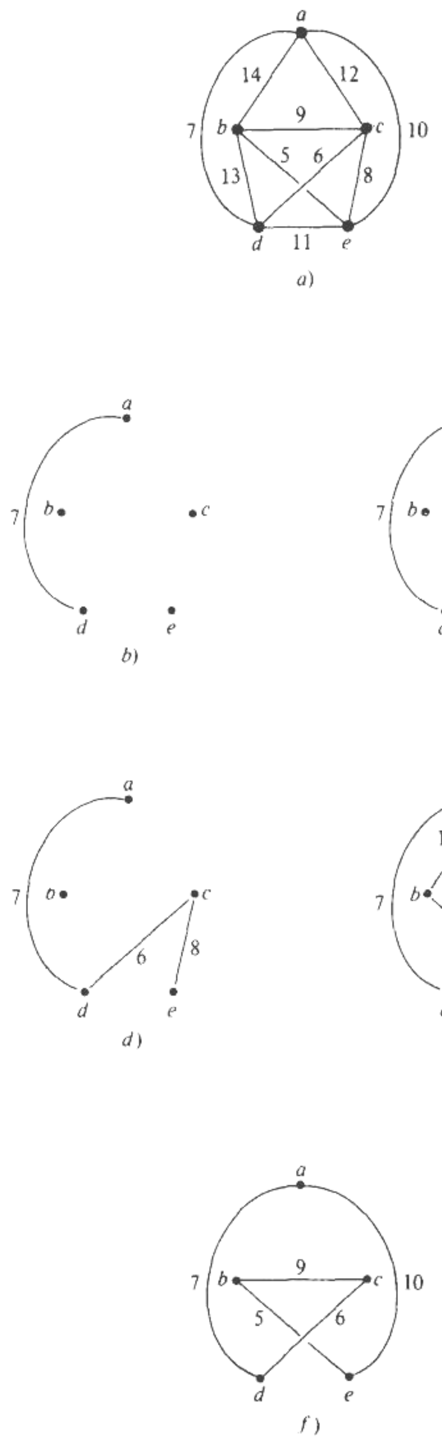

a
14
12
7
b
9
c
10
5
6
13
8
d
11
e
a)
a
7
b
c
d
e
b)
a
7
b
c
6
d
e
c)
a
7
b
c
6
8
d
e
d)
a
14
7
b
c
5
6
8
d
e
e)
a
9
7
b
c
10
5
6
d
e
f)

Figura 5.23

arista más grande en D, l_2 la longitud de la siguiente arista más grande en D, y, en general, l_1 la longitud de la i-ésima arista más grande en D. Así,

$$d = \sum_{i=1}^{n} l_i$$

Por simplicidad, tomemos $n = 14$. *Supongamos* que podemos demostrar que

$$\begin{aligned}
&d_0 \geqq 2l_1\\
&d_0 \geqq 2l_2\\
&d_0 \geqq 2(l_3 + l_4)\\
&d_0 \geqq 2(l_5 + l_6 + l_7 + l_8)\\
&d_0 \geqq 2(l_9 + l_{10} + l_{11} + l_{12} + l_{13} + l_{14})
\end{aligned} \tag{5.3}$$

entonces tendremos

$$5d_0 \geqq 2 \sum_{i=1}^{14} l_i = 2d$$

o

$$\frac{d}{d_0} \leqq \frac{5}{2} = \tfrac{1}{2}\lceil \lg 14 \rceil + \tfrac{1}{2}$$

Así podemos tener un conjunto de desigualdades similar al de las ecuaciones (5.3) para un n en general, demostraremos que

$$d_0 \geqq 2l_1 \tag{5.4}$$

$$d_0 \geqq 2 \sum_{i=k+1}^{2k} l_i \qquad 1 \leqq k \leqq \left\lceil \frac{n}{2} \right\rceil \tag{5.5}$$

y

$$d_0 \geqq 2 \sum_{i=\lfloor n/2 \rfloor + 1}^{n} l_i \tag{5.6}$$

Señalemos que si n es par (5.6) está incluido en (5.5).

Antes que nada (5.4) proviene de la desigualdad del triángulo. Supongamos que la arista más grande en D es incidente con los vértices x y y. Entonces la desigualdad del triángulo implica que la longitud de cualquier paseo entre x y y es mayor o igual a l_1. Debido a que cualquier circuito hamiltoniano de G puede dividirse en dos paseos entre x y y (5.4) se obtiene inmediatamente.

Sea a_i el vértice *para el cual* la i-ésima arista más grande en D fue agregada en nuestra construcción de acuerdo con el método del vecino más cercano (por ejemplo, según esta convención, los vértices del grafo de la figura 5.23*a* son nombrados de esta manera en la figura 5.24*a*). Para un k fijo, $1 \leqq k \leqq \lceil n/2 \rceil$, sea H el subgrafo completo de G que contiene los vértices a_i, $1 \leqq i \leqq 2k$. Sea T el circuito hamiltoniano en H que visita los vértices de H en el mismo orden circular como el de un circuito hamiltoniano mínimo

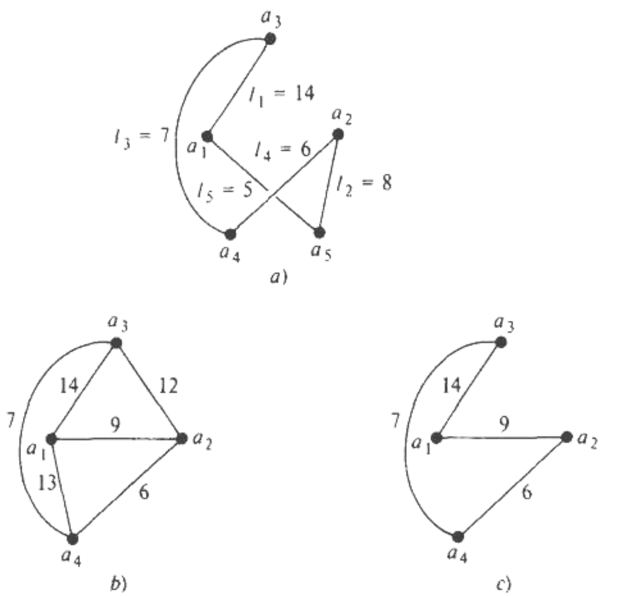


Figura 5.24

que visita los vértices de G.† Sea t la longitud de T (en el ejemplo de la figura 5.23, al considerar el caso $k = 2$. El subgrafo H es el que se muestra en la figura 5.24*b*. En el circuito hamiltoniano mínimo de la figura 5.23*f* los vértices son visitados en el circular a_2, a_1, a_5, a_3, a_4, de manera que el circuito T *es* el mostrado en la figura 5.24*c*, y $t = 36$). Por la desigualdad del triangulo, tenemos

$$d_0 \geqq t \tag{5.7}$$

Sea $\{a_i, a_j\}$ una arista de T. Si en nuestra construcción de acuerdo con el método del vecino más cercano, el vértice a, fue agregado antes que a_j, tenemos que $w(a_i, a_j) \geqq l_i$. Si a, fue agregado antes que a_i, tenemos que $w(a_j, a_i) \geqq l_j$, que es lo mismo que $w(a_i, a_j) \geqq l_j$. Así, tenemos

$$w(a_i, a_j) \geqq \text{mín}\,(l_i, l_j) \tag{5.8}$$

Sumando (5.8) para todas las aristas en T, obtenemos

$$t \geqq \sum_{(a_i, a_j) \in T} \text{mín}\,(l_i, l_j) \tag{5.9}$$

Observemos que el valor más pequeño posible para mín (l_i, l_j) en la suma en (5.9) es l_{2k} el siguiente valor más pequeño posible es l_{2k-x}, y así sucesivamente. Además, para

† Todos los vértices en G pero no en H son ignorados en el orden circular.

cualquier $1 \leqq i \leqq 2k$, l_i puede aparecer en la suma a lo más dos veces. Puesto que hay $2k$ aristas en T, la suma en (5.9) es mayor o igual que el doble de la suma de las longitudes de las k aristas más cortas. En consecuencia, tenemos

$$t \geqq \sum_{(a_i, a_j) \in T} \text{mín}\,(l_i, l_j) \geqq 2(l_{2k} + l_{2k-1} + \cdots + l_{k+1})$$
$$= 2 \sum_{i=k+1}^{2k} l_i \tag{5.10}$$

Si combinamos (5.7) y (5.10), obtenemos (5.5).

La desigualdad (5.6) puede demostrarse de manera similar:† Sea D_0 un circuito hamiltoniano mínimo. Mediante el mismo argumento usado para deducir (5.9), obtenemos

$$d_0 \geqq \sum_{(a_i, a_j) \in D_0} \text{mín}\,(l_i, l_j)$$

A partir del mismo argumento empleado para deducir (5.10), obtenemos

$$d_0 \geqq \sum_{(a_i, a_j) \in D_0} \text{mín}\,(l_i, l_j) \geqq 2(l_n + l_{n-1} + \cdots + l_{\lceil n/2 \rceil + 1}) + l_{\lceil n/2 \rceil}$$
$$\geqq 2(l_n + l_{n-1} + \cdots + l_{\lceil n/2 \rceil + 1})$$
$$= 2 \sum_{i=\lceil n/2 \rceil + 1}^{n} l_i$$

Ahora, para $k = 1, 2^1, 2^2, \ldots, 2^{\lceil \lg n \rceil - 2}$, (5.5) da lugar a las [lg n]-1 desigualdades:

$$d_0 \geqq 2l_2$$
$$d_0 \geqq 2(l_3 + l_4)$$
$$d_0 \geqq 2(l_5 + l_6 + l_7 + l_8)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$d_0 \geqq 2(l_{2^{\lceil \lg n \rceil - 2}+1} + l_{2^{\lceil \lg n \rceil - 2}+2} + \cdots + l_{2^{\lceil \lg n \rceil - 1}})^{\ddagger}$$

Si sumamos estas desigualdades, obtenemos

$$(\lceil \lg n \rceil - 1)d_0 \geqq 2\left(\sum_{i=2}^{2^{\lceil \lg n \rceil - 1}} l_i\right) \tag{5.11}$$

Puesto que

$$\left\lceil \frac{n}{2} \right\rceil + 1 \leqq 2^{\lceil \lg n \rceil - 1} + 1 \qquad \text{para } n > 1$$

(5.6) nos conduce a

† Sólo necesitamos considerar este caso cuando n es impar.
‡ Señalemos que $2^{\lceil \lg n \rceil - 1}$ es la potencia más grande de 2 que es menor que n.

$$d_0 \geqq 2 \sum_{i=2^{\lceil \lg n \rceil - 1}+1}^{n} l_i \tag{5.12}$$

Si sumamos (5.4), (5.11) y (5.12), obtenemos

$$\left(\lceil \lg n \rceil + 1\right) d_0 \geqq 2 \sum_{i=1}^{n} l_i = 2d$$

o

$$\frac{d}{d_0} \leqq \tfrac{1}{2}\lceil \lg n \rceil + \tfrac{1}{2}$$

□

*5.9 FACTORES DE UN GRAFO

La noción de un circuito hamiltoniano para un grafo puede extenderse a la de un 2-factor de un grafo. Primero presentaremos, no obstante, una definición más general: un *k-factor* de un grafo se define como un subgrafo generador del grafo cuyos grados de cada uno de sus vértices es *k*. Por ejemplo, para el grafo de la figura 5.25*a*, la figura 5.25¿ muestra un 1 -factor, y la figura 5.25*c* muestra un 2-factor del grafo. Observemos que un grafo podría tener muchos A-factores diferentes, o podría no tener ningún A-factor para algún *k*. Por ejemplo, las figuras 5.26*a* y 5.26*b* muestran dos grafos que no tienen un 1-factor, y la figura 5.26*c* muestra un grafo que no tiene un 2-factor.

Aunque podemos demostrar mediante exhaustión que el grafo de la figura 5.26*a* no

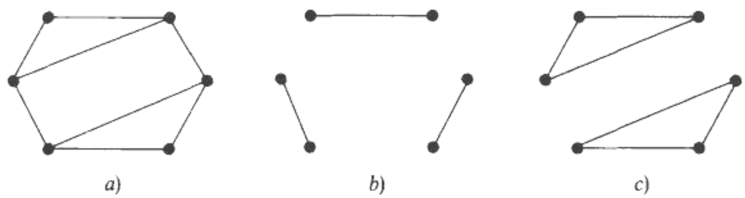


Figura 5.25

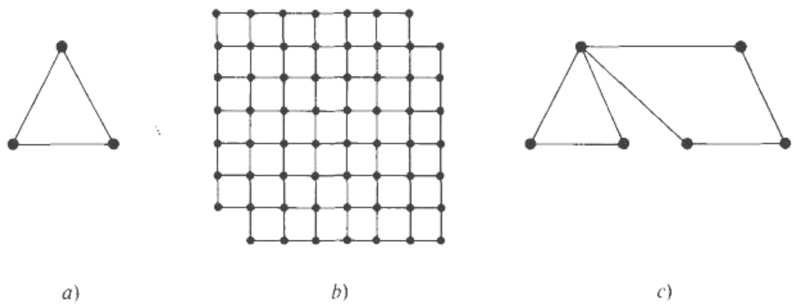


Figura 5.26

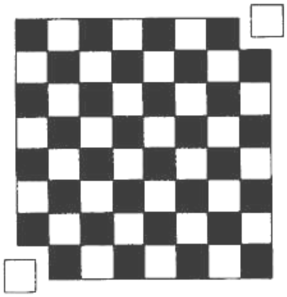


Figura 5.27

tiene un 1-factor y que el grafo de la figura 5.26*c* no tiene un 2-factor, es probable que no sea obvio que el grafo de la figura 5.*26b* no tenga un 1-factor. Ahora daremos una interesante demostración de este resultado (para otra demostración, véase el problema 5.40). Supongamos que nos dan un tablero de ajedrez con sus dos esquinas diagonalmente opuestas recortadas, como se observa en la figura 5.27. Supongamos que tenemos algunas fichas de dominó, cada una de las cuales cubre con exactitud dos de los cuadros en el tablero de ajedrez. Queremos saber si es posible cubrir el tablero truncado con 31 fichas de dominó. Señalemos que éste es exactamente el problema de determinar si hay un 1-factor en el grafo de la figura 5.266. Recordemos que los cuadros de un tablero de ajedrez estándar tienen alternados los colores blanco y negro. Puesto que los dos cuadros recortados son blancos, como se muestra en la figura 5.27, existen 30 cuadros blancos y 32 cuadros negros en el tablero de ajedrez. Debido a que una ficha de dominó siempre cubre un cuadro negro y uno blanco sin importar cómo sea colocado, se sigue que es imposible cubrir el tablero de ajedrez.

Consideremos el problema de programar reuniones para un número dado de comités. Hay dos salones de conferencias, y es posible programar dos reuniones simultáneas. No obstante, debido al traslape de las membresías de los comités, ciertos pares de reuniones no pueden ser programadas al mismo tiempo. Sea $G = (V, E)$ un grafo, donde Fes el conjunto de reuniones y una arista $\{a, b\}$ en E significa que las reuniones a y b pueden programarse al mismo tiempo. Como el lector notará, un 1-factor de G será un apareamiento aceptable.

Como se mencionó al inicio, la noción de un 2-factor es una generalización de un circuito hamiltoniano, ya que, de acuerdo con el teorema de Euler, un 2-factor es un conjunto de circuitos de vértices disjuntos tales que su unión contiene a todos los vértices del grafo. Como un ejemplo, presentaremos la solución de una adivinanza comúnmente conocida como "locura instantánea". Se nos dan cuatro cubos, con cada una de sus caras pintadas con alguno de los cuatro colores azul, verde, rojo y amarillo, como se muestra en la figura 5.28*a*, donde los cubos son numerados como 1, 2, 3, 4, y los colores son abreviados como b, g, r, y. Se nos pide apilar los cuatro cubos para formar una columna de manera que los cuatro colores se muestren en cada uno de los cuatro lados de la columna, como se ilustra en la figura 5.28*b*. En efecto, el nombre de esta adivinanza proviene de la posibilidad de que el jugador se vuelva loco al tratar de descartar el gran número de maneras diferentes para apilar los cubos. Primero observamos que los cuatro cubos pueden representarse por el grafo de la figura 5.29*a*, donde los cuatro vértices corresponden a los cuatro colores. Correspondiendo a cada cubo existen tres aristas en el grafo que especifican los colores de las tres parejas de caras opuestas. Debemos notar en este momento que, en lo que a este problema se refiere, la única informa-

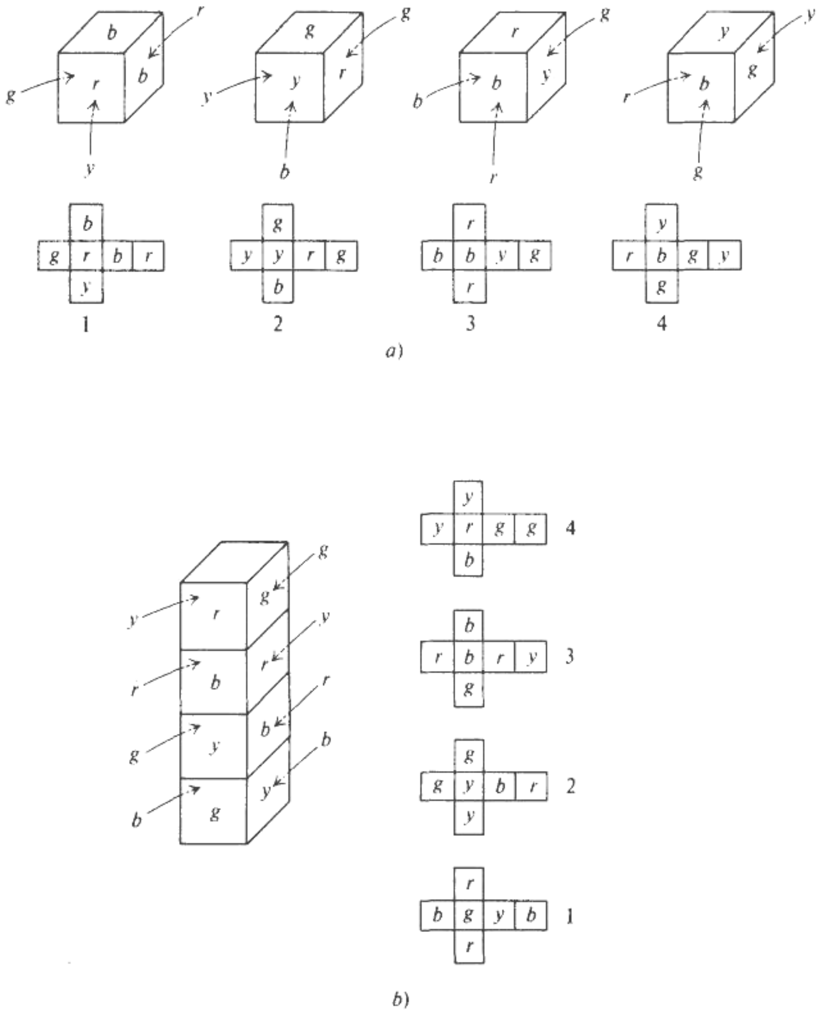


Figura 5.28

ción que necesitamos de cada cubo son los colores de las tres parejas de caras opuestas. En realidad, una lección para aprender de este ejemplo es la importancia de observar lo esencial y no confundirse con la información irrelevante cuando se trata de resolver un problema. Si buscamos una solución a este problema, notamos que los colores de los cubos sobre dos aristas opuestas de la columna pueden ser descritos mediante un 2-factor para el grafo de la figura 5.29*a*, tal que el 2-factor contiene cuatro aristas, que están etiquetadas como 1,2,3, 4, respectivamente. La figura 5.29*b* muestra un 2-factor que satisface estas condiciones. Una vez que se ha hecho esa observación, ahora la adivinanza puede replantearse como el buscar dos 2-factores de aristas disjuntas del grafo, de manera que los cuatro lados en cada 2-factor

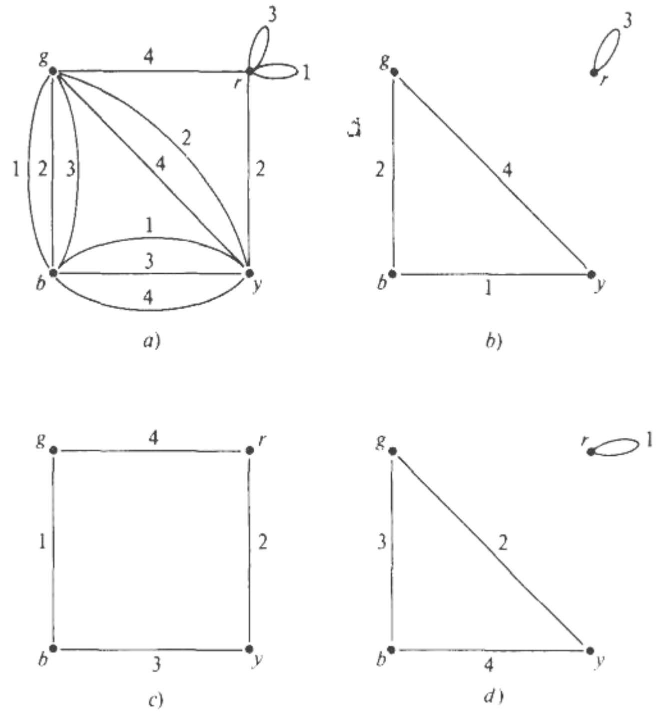


Figura 5.29

estén etiquetados en forma diferente con 1,2,3,4. Los 2-factores de las figuras 5.29*b* y 5.29*c* corresponden a la solución de la figura 5.28*b*. Después de extraer los dos 2-factores a partir del grafo de la figura 5.29*a*, se tiene la grata sorpresa al descubrir que los lados restantes, como se muestra en la figura 5.29*d*, también forman un 2-factor cuyos lados tienen etiquetas diferentes con 1, 2, 3, 4. Lo que esto significa es que nuestra solución puede tener una característica más, a saber, los cubos pueden ser orientados con los cuatro "fondos" pintados de manera distinta y las cuatro "tapas" también con pintura diferente.

*5.10 GRAFOS APLANABLES

Ahora estudiaremos una clase de grafos llamados *grafos aplanables* (o *planares).* Ésta no sólo es una clase de grafos que se encuentra con mucha frecuencia, también tiene muchas propiedades interesantes, algunas de las cuales analizaremos aquí.

Diremos que un grafo es *aplanable* si puede ser dibujado sobre un plano de manera tal que ninguna arista se cruce con otra, excepto, desde luego, en los vértices comunes. La figura 5.30*a* muestra un grafo aplanable. Observemos que el grafo de la figura 5.30*b* también es aplanable debido a que puede volverse a dibujar como se muestra en la figura 5.30*c*. La figura 5.30*d* muestra un grafo no aplanable. A decir verdad, el grafo de la figura 5.30*d* corresponde al conocido problema de determinar si es posible conectar tres casas a, *b* y c a tres servicios públicos *d, e* y *f,* de manera tal que no haya dos líneas de conexión que se

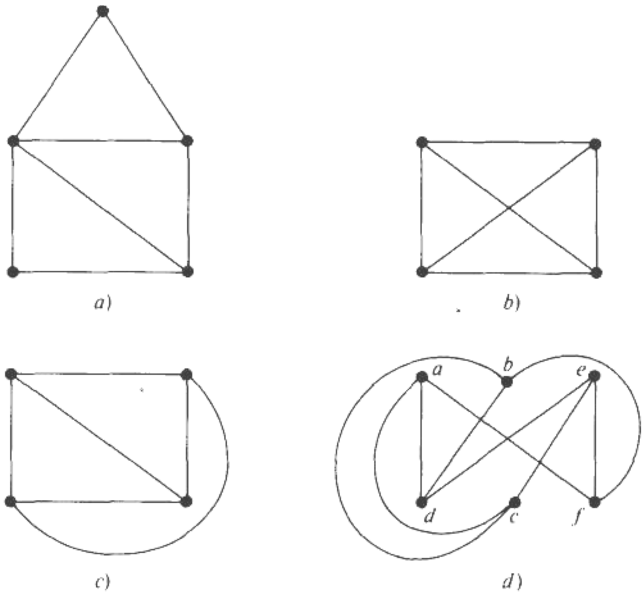


Figura 5.30

crucen una con otra. La experiencia muestra que esto no puede llevarse a cabo. En otras palabras, el grafo que representa las líneas de conexión es en realidad un grafo no aplanable (el lector puede estar en desacuerdo cuando decimos que el grafo es no aplanable sólo porque después de cierto número de intentos, encontramos que no se puede dibujar el grafo sobre un plano de manera que no se crucen las aristas. Este resultado será establecido rigurosamente más adelante).

Supongamos que dibujamos un grafo aplanable sobre el plano y tomamos un cuchillo afilado para cortar a lo largo de las aristas; entonces el plano será dividido en piezas que son llamadas las *regiones* del grafo. Para ser más formales, una región de un grafo aplanable se define como un área del plano que está acotada por aristas y no puede continuar dividiéndose en subáreas. Por ejemplo, el grafo de la figura 5.3 la tiene cinco regiones, como se muestra en la figura 5.31*b*. Observemos que si cortamos a lo largo de la arista *a* ya no dividiremos más la región 1, y si cortamos a lo largo de las aristas *b, c y d,* ya no dividiremos más la región 5. Diremos que una región *es finita* si su área es finita, y se dice que es *infinita* si su área es infinita. Es evidente que un grafo aplanable tiene exactamente una región infinita.

Tenemos el siguiente resultado:

Teorema 5.7 Para cualquier grafo aplanable conexo,

$$v - e + r = 2 \tag{5.13}$$

donde v, e y r son el número de vértices, aristas y regiones del grafo, respectivamente.

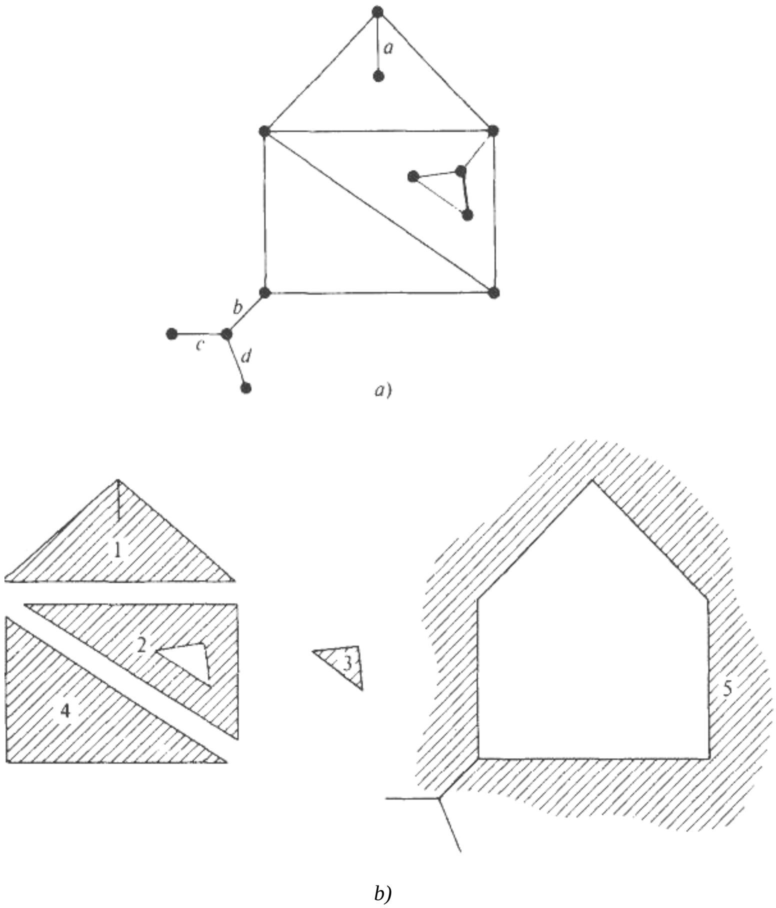


Figura 5.31

DEMOSTRACIÓN La demostración se realiza por inducción sobre el número de aristas. Como base de inducción, observemos que para los dos grafos con una sola arista mostrados en la figura 5.32, (5.13) se satisface. Como paso inductivo, supongamos que (5.13) se satisface en todos los grafos con n - 1 lados. Sea G un grafo conexo con n aristas. Si G tiene un vértice de grado 1, la eliminación de este vértice junto con la arista incidente con él, dará origen a un grafo conexo G', como se ilustra en la figura 5.33a.

Figura 5.32

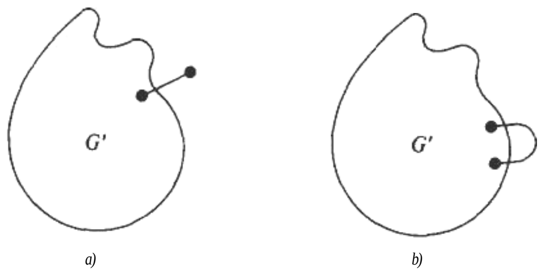


Figura 5.33

Dado que (5.13) se satisface en *G*', también se satisface en *G*, debido a que cuando se ponen de nuevo en *G* el vértice y la arista eliminados anteriormente, se incrementará el número de vértices por 1 y el número de aristas por 1, pero no habrá cambios en el número de regiones. Si *G* no tiene vértices de grado 1, al eliminar cualquier arista en la frontera de una región finita,† originaremos un grafo conexo *G*', como se ilustra en la figura 5.33*b*. Puesto que (5.13) se satisface en *G'*, también se satisface en *G*, debido a que al poner de regreso en *G'* la arista eliminada incrementaremos el número de aristas por uno y el número de regiones por uno, pero no habrá cambio en el número de vértices.‡

La ecuación (5.13) se conoce como la *fórmula de Euler* para grafos aplanables. Es verdaderamente remarcable que, sin excepción alguna, todos los grafos aplanables conexos deben satisfacer esta fórmula tan simple. Consideremos ahora algunas aplicaciones de la fórmula de Euler. Primero queremos demostrar que en cualquier grafo aplanable lineal conexo que no tenga lazos y que tenga dos o más aristas,

$$\mathbf{e} \leqq 3\mathbf{v} - 6$$

Contemos las aristas en la frontera de una región y luego calculemos el número total para todas las regiones.§ Debido a que el grafo es lineal, cada región está acotada por tres o más aristas. Por tanto, el número total es mayor o igual a 3r. Por otro lado, como una arista está en las fronteras de a lo más dos regiones, el número total es menor o igual a 2e. Así,

$$2\mathbf{e} \geqq 3\mathbf{r}$$

o

$$\frac{2\mathbf{e}}{3} \geqq \mathbf{r}$$

† No es difícil ver que, si G no tiene región finita, *G* debe contener un vértice de grado 1.

‡ (5.13) no es satisfecha por grafos aplanables no conexos. ¿Por qué nuestra demostración no es válida para grafos aplanables no conexos?

§ Se considera que aristas tales como *a, b, c, d* de la figura 5.3*a* no están en la frontera de una región cualquiera.

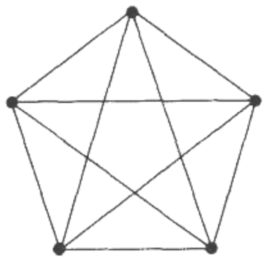


Figura 5.34

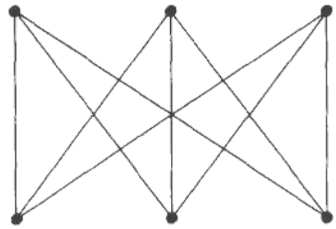


Figura 5.35

De acuerdo con la fórmula de Euler, tenemos

$$v - e + \frac{2e}{3} \geqq 2 \tag{5.14}$$

o

$$3v - 6 \geqq e \tag{5.15}$$

Ahora demostraremos que el grafo de la figura 5.34 no es un grafo aplanable. Para este grafo v = 5 y e = 10, la desigualdad en (5.15) no se satisface.

Demostraremos que el grafo de la figura 5.35 tampoco es aplanable. Si el grafo fuera aplanable cada región estaría acotada por cuatro o más aristas. Así, la desigualdad de (5.14) puede restringirse a

$$v - e + \frac{e}{2} \geqq 2$$

o

$$2v - 4 \geqq e \tag{5.16}$$

Para el grafo de la figura 5.35, v = 6 y e = 9. En consecuencia, se contradice (5.16).

A pesar de que en ocasiones la fórmula de Euler puede aplicarse para asegurar que un grafo dado no es aplanable, el argumento en dichas aplicaciones de la fórmula puede resultar complicado y engañoso, inclusive para grafos que contengan un número moderado de vértices y aristas. Además, excepto por el mapeo real de un grafo sobre el plano, no tenemos manera de asegurar que un grafo dado sea aplanable. Estableceremos un teorema, debido a Kuratowski, que nos permitirá determinar de modo inequívoco si un grafo es aplanable.

Es evidente que la planaridad de un grafo no se ve afectada si una arista es dividida en dos aristas por la inserción de un vértice de grado 2, como se muestra en la figura 5.36*a*, o si dos aristas que son incidentes con un nuevo vértice de grado 2 se combinan como una sola arista al eliminar ese vértice, como se muestra en la figura 5.36*b*. Esto sugiere la siguiente definición: diremos que dos grafos G_1 y G_2 son *isomorfos bajo vértices de grado 2* si son isomorfos o si pueden transformarse en grafos isomorfos mediante repeticiones de inserciones y/o eliminaciones de vértices de grado 2, como se muestra en la figura 5.36*a* y 5.36*b*. Por ejemplo, los dos grafos de la figura 5.36*c* son isomorfos bajo vértices de grado 2.

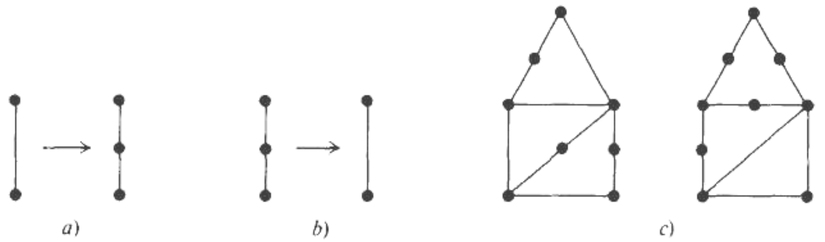


Figura 5.36

Teorema 5.8

(Kuratowski) Un grafo es aplanable si y sólo si no contiene cualquier subgrafo que sea isomorfo bajo vértices de grado 2 a cualquiera de los grafos de las figuras 5.34 o 5.35.

Los grafos de las figuras 5.34 y 5.35 también son llamados *grafos de Kuratowski.* Señalemos que el teorema de Kuratowski es efectivamente un resultado de caracterización muy significativa en el sentido de que, aunque exista multitud de grafos no aplanables, todos ellos contienen un subgrafo que es isomorfo bajo vértices de grado 2 a alguno de los grafos de Kuratowski. La demostración de este teorema es muy extensa a pesar de ser elemental. No incluimos la demostración aquí, pero puede encontrarse en Berge [3] o Liu [12].

5.11 NOTAS Y REFERENCIAS

Berge [3], Bondy [4], Harary [8] y Wilson [15] son referencias generales sobre la teoría de grafos. Consúltese también Busacker y Saaty [5] y Liu [12] para aplicaciones de la teoría de grafos. Consúltese Aho, Hopcroft y Ullman [1], Aho, Hopcroñ y Ullman [2], Even [7], Lawler [11] y Reingold, Nievergelt y Deo [13] para un estudio sobre algoritmos de grafos. Consulte Hennie [9] y Kohavi [10] para el modelo de estado finito para máquinas lógicas. El teorema 5.6 se debe a Rosenkrantz, Sterns y Lewis [14].

1. Aho, A. V., J. E. Hopcroft y J. D. Ullman: *The Design and Analysis of Computer Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1974.
2. Aho, A. V., J. E. Hopcroft y J. D. Ullman: *Data Structures and Algorithms,* Addison-Wesley Publishing Company, Reading, Mass, 1983.
3. Berge, C: *The Theory of Graphs and Its Applications,* John Wiley & Sons, Nueva York, 1962.
4. Bondy, J. A. y U. S. R. Murty: *Graphs Theory with Applications,* American Elsevier Publishing Company, Nueva York, 1976.
5. Busacker, R. G. y T. L. Saaty: *Finite Graphs and Networks: An Introduction with Applications,* McGraw-Hill Book Company, Nueva York, 1965.
6. Dijkstra, E. W.: "ANote on Two Problems in Connexion with Graphs", *Numerische Mathematik,* 1: 269-271 (1959).
7. Even, S.: *Graph Algorithms,* Computer Science Press, Potomac, Md., 1979.
8. Harary, R: *Graph Theory,* Addison-Wesley Publishing Company, Reading, Mass., 1969.
9. Hennie, R C: *Finite-State Models for Logical Machines,* John Wiley & Sons, Nueva York, 1968.

10. Kohavi, Z.: *Switching and Autómata Theory,* 2ª ed, McGraw-Hill Book Company, Nueva York, 1978.
11. Lawler, E: *Combinatorial Optimization,* Holt, Rinehart and Winston, Nueva York, 1976.
12. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
13. Reingold, E. M., J. Nievergelt y N. Deo: *Combinatorial Algorithms: Theory and Practice,* Prentice-Hall, Englewood Cliffs, N. J., 1977.
14. Rosenkrantz, D. J., R. E. Stearns y P. M. Lewis: "An Analysis of Several Heuristics for the Traveling Salesperson Problems", *SIAM Journal on Computing,* 6: 566-581 (1977).
15. Wilson, R. J.: *Introduction to Graph Theory,* Academic Press, Nueva York, 1972.

PROBLEMAS

5.1 Se espera que un hombre traslade un perro, una oveja y un costal de col a través de un río por medio de una canoa. La canoa es muy pequeña, y él puede llevar sólo uno de estos tres objetos a la vez. Además, él no puede dejar al perro solo con la oveja ni a la oveja sola con la col. Con el objeto de determinar cómo debe proceder, descríbanse los pasos que debe seguir mediante un grafo no dirigido. Considérese que los vértices del grafo representan todas las configuraciones permitidas. Por ejemplo: el hombre, el perro, la oveja y la col están todos a un lado del río en la configuración inicial; el hombre, el perro, la oveja y la col están todos del otro lado del río en la configuración final. Elhombre, el perro y la oveja están de un lado del río en tanto la col está al otro lado del río, es una configuración intermedia permitida. Habrá una arista entre dos vértices si el hombre puede realizar un viaje a través del río de manera que la configuración correspon diente a uno de los vértices pueda transformarse en la correspondiente al otro vértice, y recíprocamente. Construya el grafo y determine todas las posibles maneras en que el hombre transporta los objetos a través del río.

5.2 *a)* Durante un viaje tres parejas de casados llegan a un río donde encuentran un bote que no puede transportar a más de dos personas a la vez. El paso a través del río se ve complicado por el hecho de que todos los esposos son muy celosos y no permiten a sus esposas quedarse sin ellos cuando están presentes otros hombres. Construya un grafo que muestre cómo puede efectuarse el traslado.

b) Demuestre que el problema del inciso *a)* no puede solucionarse si hay cuatro parejas.

c) Demuestre que el problema del inciso *a)* puede solucionarse si hay cuatro parejas y el bote puede llevar tres personas.

5.3 Un circuito electrónico se construye para reconocer sucesiones de números ceros y unos. En particular aceptará sucesiones de la forma 010* 10, donde 0* significa cualquier número (inclusive ninguno) de ceros. Por ejemplo, 0110, 01010 y 010000010 son todas sucesiones aceptables. Construya un grafo dirigido en el cual cada vértice tenga dos aristas de salida etiquetadas 0 y 1, y en el que existan dos vértices, v_i (el vértice inicial) y v_f (el vértice final), tales que cualquier paseo desde v_i hasta v_f sea una sucesión de la forma 010*10 (para cada sucesión el circuito iniciará en v_i y recorrerá las aristas de acuerdo con la sucesión de números 0 y 1. El circuito aceptará la sucesión si el paseo termina en v_f).

5.4 Repita el problema 5.3 para sucesiones de la forma 01*(10)* 10*, donde (10)* significa cualquier número (inclusive ninguno) de patrones 10, y use cada una de las siguientes modificaciones:

a) Para cada vértice no hay restricción sobre el número de aristas salientes etiquetadas con 0 o 1. La sucesión de etiquetas en cada paseo de v_i a v_f es una sucesión de la forma descrita. Además, para cada sucesión de la forma ya descrita existe un paseo correspondiente de v_i a v_f

b) Podrá haber varios vértices finales $v_{f_1}, v_{f_2}, \ldots$ La sucesión de etiquetas de cada paseo desde v_i hasta algún f_f final es una sucesión de la forma anteriormente descrita. Además, para cualquier sucesión de la forma ya descrita, existe un paseo correspondiente desde v_i hasta algún vértice final f_i.

Reunión	Hora de inicio
Primera	8:00 a.m.
Segunda	9:00 a.m.
Tercera	11:00 a.m.
Cuarta	2:00 p.m.
Quinta	3:30 p.m.
Sexta	4:00 p.m.

d)

Reunión	Duración estimada de la reunión
Primera	90 min.
Segunda	60 min.
Tercera	120 min.
Cuarta	120 min.
Quinta	90 min.
Sexta	30 min.

b)

Figura 5P.1

5.5 El jefe del departamento de ciencias de la computación quiere convocar seis reuniones de comité en un cierto día. Después de enviar las notificaciones de los tiempos de inicio para las reuniones, como se muestra en la figura 5P.1*a*, su secretaria, la Srita. Masón, descubrió que las agendas de estas reuniones habían sido revisadas. En consecuencia, ella realizó una nueva estimación de la duración de estas reuniones, como se muestra en la figura 5P.1*b*. Dado que las reuniones deben realizarse en el orden en que fueron programadas originalmente, fue necesario realizar cambios en la hora de inicio de algunas de las reuniones. Con la esperanza de efectuar el menor número de cambios posible, la Srita. Masón elaboró el siguiente grafo no dirigido: sean v_1, v_2,..., v_6 seis vértices que representan las seis reuniones. Existe una arista entre v_i y v_j para $i < 1$ si la suma de las duraciones de la i-ésima, la $(i + 1)$-ésima, la $(i + 2)$-ésima,.. ., y la $(j$ - l)-ésima reuniones es menor o igual que la diferencia entre los tiempos de inicio originalmente escogidos para la i-ésima y j-ésima reuniones.

a)¿Cuál es el significado físico de la arista entre v_i y v,?, ¿y del subgrafo completo que contiene a $v_{i_1}, v_{i_2}, \ldots, v_{i_k}$?, ¿del subgrafo conexo más grande posible?

b) Suponga que en la reprogramación de las reuniones, la Srita. Masón quiere asegurarse de que la primera reunión no puede comenzar antes de las 8:00 a.m., y que la última reunión no se prolongará más allá de las 5:00 p.m. ¿Cómo puede hacer uso de un programa de computadora capaz de determinar el subgrafo completo más grande posible para un grafo dado? Determine un nuevo conjunto de horas para el inicio de las reuniones con el menor número posible de cambios de los originales.

5.6 Sea $G = \{V,E)$ un grafo no dirigido con k componentes y $|V| = n$, $|E| = m$. Demuestre que $m \geqq n - k$.

5.7 n ciudades están comunicadas por una red de k autopistas (una autopista se define como una carretera entre dos ciudades que no pasa a través de ciudades intermedias). Demuestre que si $k > \frac{1}{2}(n-1)(n-2)$, entonces uno siempre puede viajar entre dos ciudades cualesquiera a través de autopistas que les comuniquen.

5.8 Diremos que una «-ada ordenada $(d_u\ d_2,..., d_n)$ de enteros no negativos es *graficable* si existe un grafo lineal, sin lazos, que tenga n vértices, cuyos grados de sus vértices son $d_u\ d_2,..., d_n$.

a) Demuestre que (4, 3, 2, 2,1) es graficable.

b) Demuestre que (3, 3, 3,1) no es graficable.

c) Sin pérdida de generalidad, suponga que $d_1 \geqq d_2 \geqq d_3 \geqq \cdots \geqq d_n$. Demuestre que $(d_1\ d_2, \ldots, d_n)$ es graficable si y sólo si $(d_2 - 1, d_3 - 1, \ldots, d_{d_1} - 1, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n)$ es graficable.

d) Use el resultado del inciso c) para determinar si (5, 5, 3, 3, 2, 2, 2) es graficable.

5.9 *a)* Demuestre que la suma de los grados de entrada de todos los vértices es igual a la suma de los grados de salida de todos los vértices en cualquier grafo dirigido.

b) Demuestre que la suma de los cuadrados de los grados de entrada de todos los vértices es igual a la suma de los cuadrados de los grados de salida de todos los vértices en cualquier grafo completo dirigido.

5.10 Diremos que un grafo es *autocomplementario* si es isomorfo a su complemento.

a) Muestre un grafo autocomplementario con cuatro vértices.

b) Muestre un grafo autocomplementario con cinco vértices.

c) ¿Existe un grafo autocomplementario con tres vértices?, ¿con seis vértices?

d) Demuestre que un grafo autocomplementario debe tener $4k$ o $4k + 1$ vértices.

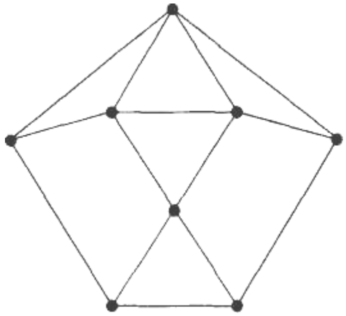


Figura 5P.2

5.11 Diremos que un conjunto de vértices en un grafo no dirigido es un *conjunto dominante* si para todo vértice que no está en el conjunto éste es adyacente a uno o más vértices en el conjunto. Un *conjunto dominante minimal* es un conjunto dominante tal que no existe un subconjunto propio de éste que sea también un conjunto dominante.

a) Para el grafo de la figura 5P.2 encuentre dos conjuntos dominantes minimales de diferentes tamaños.

b) Si los vértices de un grafo representan ciudades y las aristas, vías de comunicación entre dos ciudades dé una interpretación física de la noción de conjunto dominante de este caso.

c) Considere que los 64 cuadros de un tablero de ajedrez están representados por 64 vértices, y que existe una arista entre dos vértices si los cuadros correspondientes están en la misma fila, misma columna, o misma diagonal (hacia atrás o hacia adelante). Se sabe que cinco reinas pueden colocarse en un tablero de ajedrez de manera que dominen los 64 cuadros. Además, cinco es el número mínimo de reinas necesarias. Plantee este enunciado en términos de teoría de grafos.

5.12 Diremos que un conjunto de vértices en un grafo no dirigido es un *conjunto independiente* si no existen dos vértices en este conjunto que sean adyacentes. Un *conjunto independiente maximal* es un conjunto independiente el cual no continuará siéndolo cuando le sea agregado un vértice cualquiera.

a) Para el grafo de la figura 5P.2 encuentre dos conjuntos independientes maximales de tamaños diferentes.

b) ¿Cómo puede enunciarse en términos de teoría de grafos el problema de colocar ocho reinas en un tablero de ajedrez de forma tal que ninguna de ellas capture a otra?

5.13 *a)* Las aristas de un K_6 se pintan de rojo o azul. Demuestre que para cualquier manera arbitraria de pintar las aristas habrá ya sea un K_3 rojo (un K_3 con todas sus aristas pintadas de rojo) o bien un K_3 azul.

b) Utilice el resultado del inciso *a)* para demostrar que de entre un grupo de seis personas existen tres de ellos que son amigos o bien tres de ellos que son extraños uno al otro.

c) Las aristas de K_n se pintan de rojo o azul de manera arbitraria. Demuestre que si existen seis o más aristas rojas que incidan con un vértice, entonces existe un K_4 rojo o bien un K_3 azul. Demuestre que si existen cuatro o más aristas azules incidiendo con un vértice, entonces existe bien un K_4 rojo o bien un K_3 azul.

d) Demuestre que para cualquier manera arbitraria de pintar los lados de K_9 en rojo o azul, existe bien un K_4 rojo o un K_3 azul.

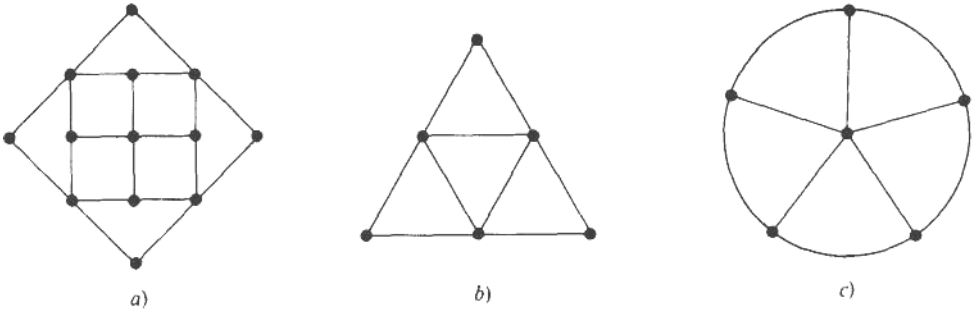


Figura 5P.3

5.14 Por *colorido apropiado* de un grafo entendemos pintar los vértices del grafo con uno o más colores distintos de manera que no hay dos vértices adyacentes pintados con el mismo color.

a) ¿Cuál es el número mínimo de colores necesarios para colorear apropiadamente el grafo de la figura 5P.3*a*?, ¿el grafo de la figura 5P.3*b*?, ¿el grafo de la figura 5P.3*c*?

b) Sea *G* un grafo lineal sin lazos. Considere que los vértices de *G* representan los exámenes a ser aplicados durante el periodo de evaluación final, y las aristas de *G* representan restricciones tales como que, una arista entre dos vértices significa que los exámenes correspondientes no se pueden programar al mismo tiempo. ¿Qué interpretación puede darse al colorido apropiado de G?, ¿al número mínimo de colores necesarios para el colorido apropiado de G?

5.15 *a)* Sea $G = (V, E)$ un grafo dirigido lineal donde *V* representa un conjunto de personas y *E* representa una relación padre-hijo tal que una arista (a, b) en *E* significa que *a* es padre de *b*. *G* recibe el nombre de un *grafo genético.* Observamos que *G* tiene las siguientes propiedades:

1. El grado de incidencia de cualquier vértice es a lo más 2.
2. No existe un circuito en *G*.
3. Los vértices de *G* pueden colorearse con dos colores de manera que para dos aristas cua lesquiera $\{a, c)$ y (b, c) en *E*, *a* y *b* están coloreados con diferente color.

¿Qué interpretación genética puede dar a estas condiciones?

b) Sea $\hat{G} = (V, \hat{E})$ un grafo no dirigido tal que existe una arista $\{a, b)$ en $\hat{E}$ si y sólo si existen dos aristas $\{a, c)$ y (b, c) en *E* para algún c. Demostrar que la condición (3) puede satisfacerse si y sólo si $\hat{G}$ puede ser coloreado apropiadamente con dos colores (véase el problema 5.14 para la definición de colorido apropiado).

c) Demuestre que un grafo no dirigido puede ser de colorido apropiado con dos colores si y sólo si éste no contiene circuitos de longitud impar.

d) Definimos un circuito alternativo en un grafo dirigido como una sucesión de aristas (v_1, v_2) (v_3, v_2) (v_3, v_4) (v_5, v_4) $\cdots$ (v_{k-1}, v_k) (v_1, v_k) tal que al invertir la dirección de las aristas alternantes de la sucesión daremos origen a un circuito (dirigido). Demuestre que la condición (3) puede satisfacerse si y sólo si la longitud de cualquier circuito alternante en *G* es divisible por 4.

5.16 La *distancia* entre dos vértices en un grafo no dirigido se define como el número mínimo de aristas entre ellos. El *diámetro* de un grafo no dirigido se define como el máximo de las distancias de las aristas entre todos los pares de vértices.

a) ¿Cuál es el diámetro del grafo en la figura 5P.4?

b) Considere que los vértices de un grafo representan computadoras y las aristas representan líneas de comunicación de datos entre las computadoras. En este caso, ¿cuál es el significado físico del diámetro del grafo?

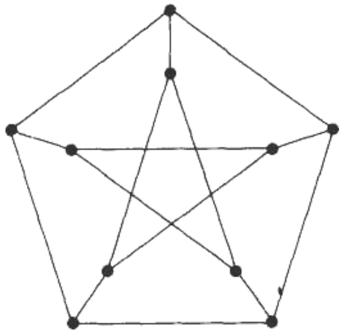


Figura 5P.4

c) Si *d* denota el diámetro de un grafo con *n* vértices, y si 5 denota el máximo de los grados de los vértices, demuestre que

$$1 + \delta + \delta(\delta - 1) + \delta(\delta - 1)^2 + \cdots + \delta(\delta - 1)^{d-1} \geqq n$$

[Así, si queremos diseñar una red de comunicaciones de computadoras con las restricciones: 1) una computadora puede comunicarse con cualquier otra computadora sin pasar a través de más de *d* - 1 computadoras intermedias; 2) una computadora puede comunicarse directamente con a lo más 8 computadoras, entonces no podemos tener una red que contenga "demasiadas" computadoras.]

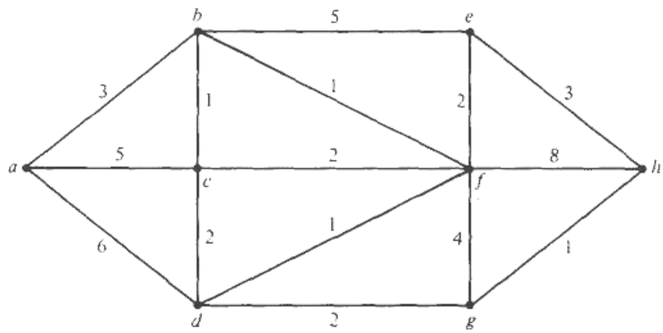


Figura 5P.5

5.17 El grafo de la figura 5P.5 muestra los canales y los tiempos de retraso en los canales de comunicación entre ocho centros de comunicación. Los centros son representados por vértices, los canales por aristas, y los tiempos de retraso (en minutos) en cada canal están representados por el peso de la arista. Suponga que a las 3:00 p.m. un centro de comunicaciones *a* transmite a través de todos sus canales la noticia de que alguien ha encontrado una manera de construir una mejor trampa para ratones. Otros centros de comunicación transmitirán esta noticia a través de

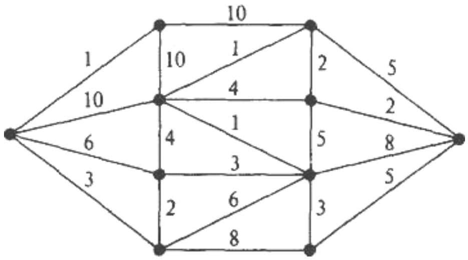


Figura 5P.6

todos sus canales tan pronto como la reciban. Para los centros de comunicación *b,c,d,e,f,g*y*h* determine el tiempo más corto en que cada uno recibe la noticia.

5.18 Aplique el algoritmo presentado en la sección 5.5 para determinar el paseo mas corto entre *a y z* en el grafo de la figura 5P.6, donde los números asociados con las aristas son las distancias entre vértices.

5.19 Aplique el algoritmo presentado en la sección 5.5 para determinar el paseo más corto entre *a y z* en el grafo de la figura 5P.7, donde los números asociados con las aristas son las distancias entre vértices.

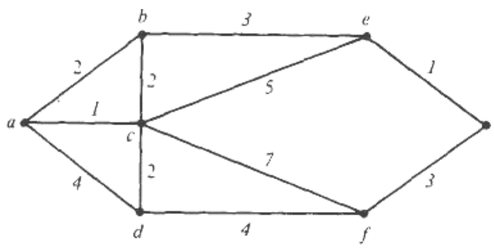


Figura 5P.7

5.20 Aplique el algoritmo presentado en la sección 5.5 para determinar el paseo más corto entre *a y z* en el grafo de la figura 5P.8, donde los números asociados con las aristas son las distancias entre vértices.

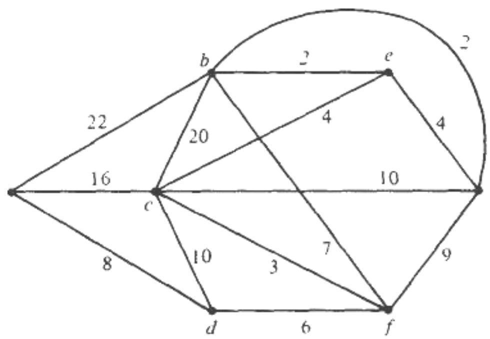


Figura 5P.8

5.21 Aplique el algoritmo presentado en la sección 5.5 para determinar el paseo más corto entre *a* y *z* en el grafo de la figura 5P.9, donde los números asociados con las aristas son las distancias entre vértices.

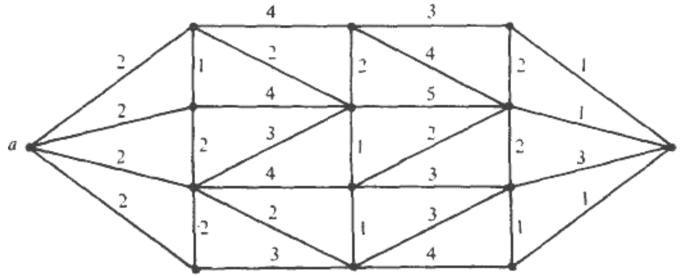


Figura 5P.9

5.22 Aplique el algoritmo de la sección 5.5 para determinar el paseo más corto entre *a* y *z* en el grafo de la figura 5P.10, donde los números asociados con las aristas son las distancias entre vértices.

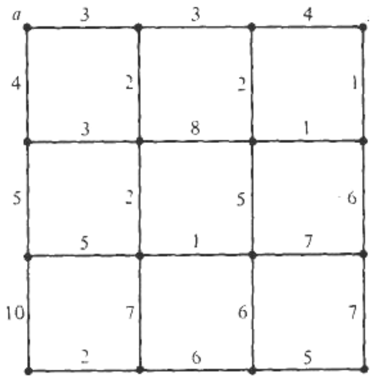


Figura 5P.10

5.23 Sea *G* un grafo conexo sin lazos donde las aristas representan las calles de una ciudad. Un oficial de policía quiere hacer un recorrido para patrullar cada lado de cada calle exactamente una vez. Además, el oficial quiere patrullar los dos lados de una calle en direcciones opuestas. Demuestre que tal recorrido siempre puede diseñarse.

5.24 Entre las muchas habitaciones de una vieja mansión existe un fantasma en cada habitación, que tiene un número par de puertas. Si la mansión tiene sólo una entrada, demuestre que una persona que entra desde el exterior siempre podrá llegar a una habitación en la cual no hay un fantasma.

5.25 *a)* Las aristas del grafo de la figura 5P.11 pueden particionarse en dos paseos (de aristas disjuntas). Muestre una de tales particiones.

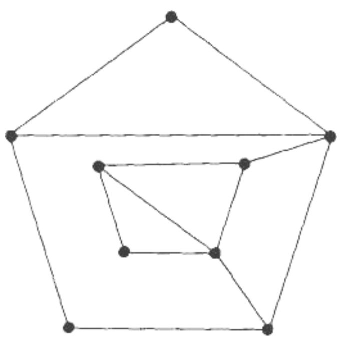


Figura 5P.11

b) Sea *G* un grafo conexo con *k* vértices de grado impar ($k > 0$). Demuestre que las aristas de *G* pueden particionarse en *k/2* paseos (de aristas disjuntas).

c) Sea *G* un grafo con *k* vértices de grado impar ($k > 0$). ¿Cuál es el número mínimo de aristas que puede agregarse a G de manera que el grafo resultante tenga un circuito euleriano? Muestre cómo se puede hacer esto para el grafo de la figura 5P. 11. Establezca cómo se puede hacer esto en general.

d) En el inciso c), suponga que sólo se nos permite agregar aristas que sean paralelos a las aristas existentes de G. ¿Cuál es el número mínimo de aristas que puede agregarse de manera que el grafo resultante tenga un circuito euleriano? ¿Se puede hacer esto siempre? Establezca una condición necesaria y suficiente bajo la cual esto puede hacerse.

5.26 ¿Es posible mover un caballo sobre un tablero de ajedrez de 8 x 8 de manera que complete cualquier movimiento posible exactamente una vez? Un movimiento entre dos cuadros del tablero de ajedrez se completa cuando dicho movimiento se realiza en cualquier dirección.

5.27 Encuentre un arreglo circular para nueve letras *a*, nueve letras *b* y nueve letras *c* tal que cada una de las 27 palabras de longitud 3 a partir del alfabeto *{a, b,c)* aparezcan exactamente una vez.

5.28 *a)* Muestre un grafo que tenga tanto un circuito euleriano como un circuito hamiltoniano.

b) Muestre un grafo que tenga un circuito euleriano pero que no tenga un circuito hamiltoniano.
c) Muestre un grafo que no tenga un circuito euleriano pero que sí tenga un circuito hamilto niano.
d) Muestre un grafo que no tenga un circuito euleriano ni tampoco un circuito hamiltoniano.

5.29 *a)* ¿Tiene K_{13} un circuito euleriano?, ¿un circuito hamiltoniano?
b) Repita el inciso a) para K_{14}.

5.30 Un grafo *bipartido completo* K_{mn} es un grafo con $V= V_1 \cup V_2$ donde el conjunto de vértices es tal que no existen aristas que unan cualesquiera dos vértices de V_l o cualesquiera dos vértices en V_2, pero existe una arista que une cualquier vértice de F, con todos los vértices de V_2.

a) ¿Existe un circuito hamiltoniano en $K_{4\ 4}$?, ¿en $K_{4\ 5}$?, ¿en $K_{4\ 6}$?
b) ¿Existe un paseo hamiltoniano en $K_{4,\ 4}$?, ¿en $K_{4\ 5}$?, ¿en $K_{4\ 6}$?
c) Establezca una condición necesaria y suficiente para la existencia de un circuito hamiltoniano en $K_{m<n}$.
d) Establezca una condición necesaria y suficiente para la existencia de un paseo hamiltoniano en $K_{m¡n}$.

5.31 Demuestre que el grafo de la figura 5P. 12 no tiene un circuito hamiltoniano.

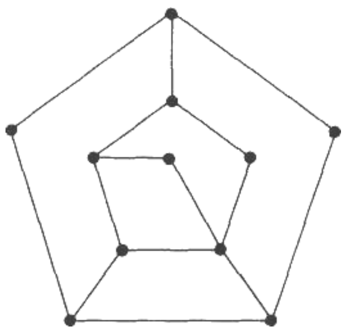


Figura 5P.12

5.32 Demuestre que cualquier circuito hamiltoniano en el grafo mostrado en la figura 5P.13 que contiene la arista *x* deberá contener también la arista *y*.

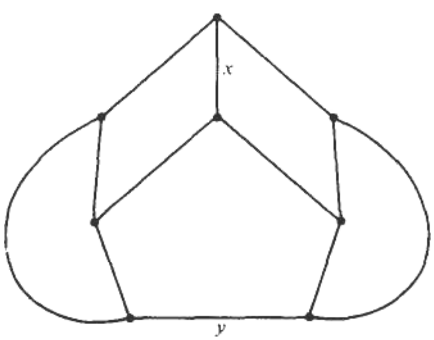


Figura 5P.13

5.33 Sea G un grafo no dirigido lineal con n vértices, $n \geqq 3$. Si los vértices de G representan n personas y las aristas de G representan una relación de amistad entre ellas tal que dos vértices están conectados por una arista si y sólo si las personas correspondientes son amigos.

a) ¿Qué interpretación puede darse al grado de un vértice?

b) ¿Qué interpretación puede darse al hecho de que G sea un grafo conexo?

c) ¿Qué interpretación puede darse a un subgrafo de G que es un grafo completo con m vértices?

d) ¿Qué interpretación puede darse a un 1-factor de G?

e) Suponga que entre dos personas cualquiera de ellas conoce a todas las $n - 2$ personas restantes. Demuestre que las n personas pueden ser sentadas en una línea de manera que cada una permanezca junto a dos de sus amigos, excepto las dos personas que están en los dos extremos de la línea, cada una de ellas permanece junto a sólo uno de sus amigos.

Sugerencia: aplique el teorema 5.4.

f) Extienda el teorema 5.4 para demostrar que si la suma de los grados de cada par de vértices de un grafo lineal con n vértices es mayor o igual a n, entonces existe un circuito hamiltoniano en G.

g) Utilice el resultado del inciso *f*) para demostrar que la condición en el inciso *e)* garantiza que las n personas pueden sentarse alrededor de un círculo de manera que cada una permanece junto a dos de sus amigos para $n \geqq 4$.

5.34 Sea G un grafo dirigido completo. Diremos que un subconjunto no vacío de los vértices de G es un *grupo fuera de clase* si cualquier arista que una un vértice en el subconjunto y un vértice fuera del subconjunto siempre está dirigido desde el último hacia el primero. Demuestre que G tiene un circuito dirigido que contiene todos los vértices si no existe un grupo de vértices fuera de clase.

5.35 Once estudiantes planean salir a cenar juntos durante varios días. Ellos se sentarán en una mesa redonda, y el plan consiste en que cada estudiante tendrá diferentes vecinos en cada cena. ¿Durante cuántos días podrá hacerse esto?

5.36 Un *n-cubo* es un grafo no dirigido con 2" vértices los cuales están etiquetados con los 2" números binarios de *n*-dígitos. Existe una arista entre dos vértices si sus etiquetas binarias difieren exactamente en un dígito. Demuestre que un n-cubo tiene un circuito hamiltoniano para cualquier $n \geqq 1$ (un arreglo secuencial de los 2*n* números binarios de n-dígitos tal que cada dos números adyacentes difieran en exactamente un dígito se conoce como un *código Gray).*

5.37 Diremos que un grafo no dirigido es *orientable* si se puede asignar una dirección a cada una de las aristas del grafo de manera que el grafo dirigido resultante sea fuertemente conexo.

a) Demuestre que el grafo de la figura 5P.14 es orientable.

b) Demuestre que cualquier grafo con un circuito euleriano es orientable.

c) Demuestre que cualquier grafo con un circuito hamiltoniano es orientable.

d) Demuestre que un grafo conexo no dirigido es orientable si y sólo si cada arista del grafo está contenida en al menos un circuito.

Figura 5P.14

5.38 *a*) Utilice el método del vecino más cercano para determinar un circuito hamiltoniano para el grafo de la figura 5P.15, a partir del vértice *a*.

b) Repita el inciso *a),* pero ahora comience en el vértice *d.*

c) Determine un circuito hamiltoniano mínimo para el grafo de la figura **5P. 15.**

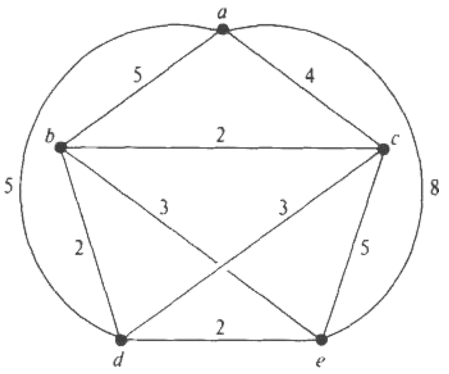


Figura 5P.10

5.39 En este problema presentamos un procedimiento que produce un buen resultado, aun cuando no el mejor posible, para el problema del agente viajero analizado en la sección 5.8. De nuevo, con sideramos que $G = (V, E, w)$ es un grafo completo de *n* vértices, donde *w* es una función de *E* al conjunto de números reales positivos que satisface la desigualdad del triángulo. Sea *T* un árbol generador mínimo de *G* (véase la sección 6.7 para la definición de un árbol extendido mínimo).

a) Sea *V* el subconjunto de vértices que tienen grados impares en el subgrafo *T.* Aseguramos que debe existir un número par de vértices en *V.* ¿Por qué?

b) Sea *G*' el subgrafo completo de *G* que contiene todos los vértices de *V.* Mdenotaun 1-factor de *G'* que tiene peso mínimo (el peso de un 1-factor se define como la suma de los pesos de las aristas en el 1-factor). Sea $T \cup M$el grafo obtenido al sobreponer los grafos *T* y *M*(si una arista aparece tanto en *T* como *M,* éste aparecerá dos veces en $T \cup M$). Demuestre que $T \cup M$ tiene un circuito euleriano.

c) Demuestre que el peso de *Tes* menor que el de un circuito hamiltoniano mínimo.

d) Demuestre que el peso de *M* es menor o igual a la mitad del circuito hamiltoniano mínimo.

e) Muestre cómo podemos obtener un circuito hamiltoniano de *G* a partir del circuito euleriano de$T \cup M$, evitando visitar los vértices que ya han sido visitados, de tal manera que el peso del circuito hamiltoniano sea menor o igual al del circuito euleriano. Consecuentemente, concluimos que el peso del circuito hamiltoniano obtenido es menor que 1.5 veces el de un circuito hamiltoniano mínimo.

Por ejemplo, para el grafo pesado de la figura 5P.16a, los pesos de todas las aristas son 1, excepto el peso de la arista {*f, g}* el cual es 2. La figura 5P. 166 muestra un árbol extendido mínimo *T.* La figura 5P.16c muestra un 1-factor del subgrafo completo *G'* el cual contiene los vértices *{a, b, c, e,f, g}.* La figura 5P.16*d* ¿ muestra un circuito euleriano de $T \cup M$, y la figura 5P.16*e* muestra un circuito hamiltoniano obtenido a partir del circuito euleriano de la figura 5P.16*j*, donde el circuito euleriano es recorrido de acuerdo con la sucesión de vértices *(a, b, c, d, e, f, d, a, g, a).*

5.40 En este problema mostramos una demostración alternativa para el resultado de la imposibilidad de cubrir un tablero de ajedrez truncado con fichas de dominó, presentado en la sección 5.9. Identifique cada cuadro del tablero de ajedrez mediante el término x^iy^j, $0 \leqq i \leqq 7$ y $0 \leqq j \leqq 7$, donde *i* es la coordenada *x* del cuadro y *j* es la coordenada *y* del cuadro. Así, la suma de los 64 términos correspondientes a los 64 cuadros en un tablero de ajedrez es

$$(1 + x + x^2 + \cdots + x^7)(1 + y + y^2 + \cdots + y^7)$$

y la suma de los 62 términos correspondientes a los 62 cuadros en el tablero de ajedrez truncado es

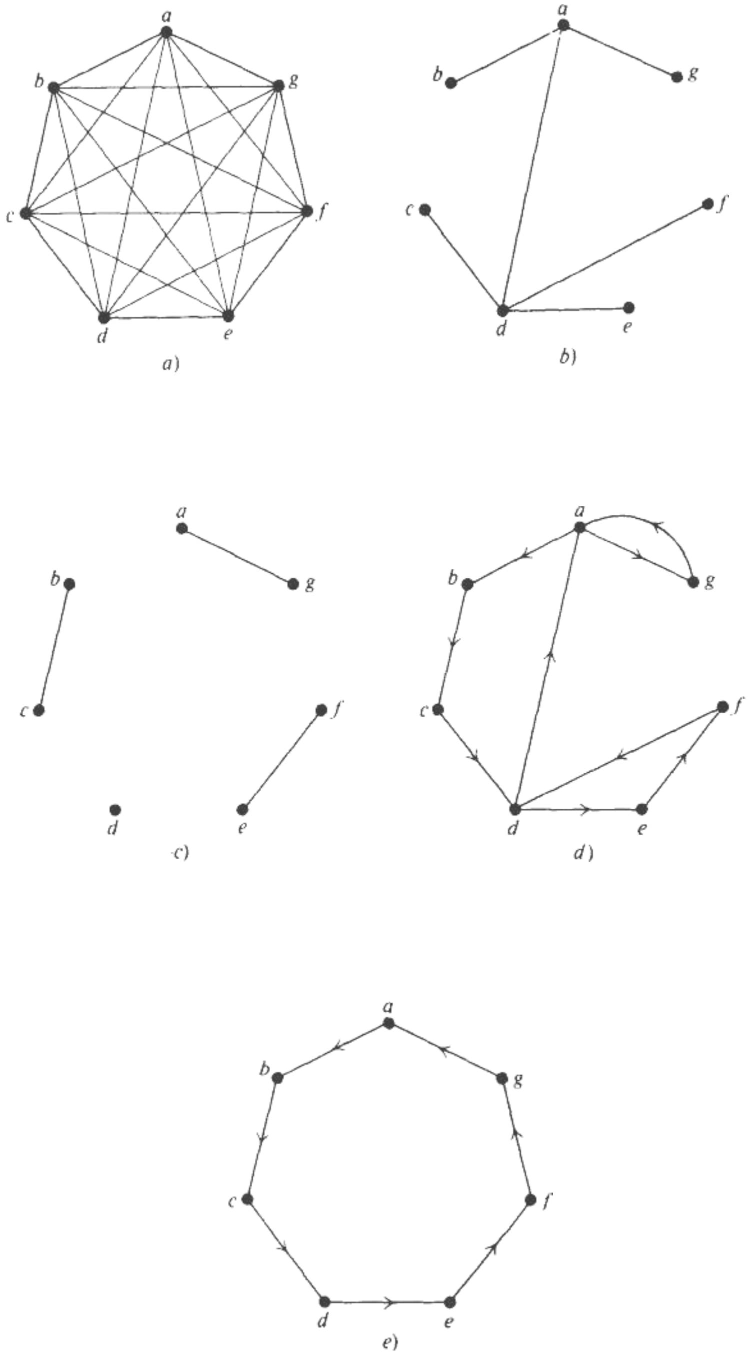


Figura 5P.16

$$(1 + x + x^2 + \cdots + x^7)(1 + y + y^2 + \cdots + y^7) - 1 - x^7y^7 \qquad \text{(5P.1)}$$

Suponga que el tablero de ajedrez truncado puede cubrirse mediante 31 fichas de dominó. Observe que una ficha de dominó colocada horizontalmente cubre dos cuadros cuyos términos correspondientes son x^iy^j y $x^{i+1}y^j$. Por tanto, la suma de los términos correspondientes a todos los cuadros cubiertos al colocar horizontalmente las fichas de dominó es de la forma $(1 + x)f(x, y)$. De modo similar, la suma de los términos correspondientes a todos los cuadros cubiertos al colocar verticalmente las fichas de dominó es de la forma $(1 + y)g(x, y)$. Así, la suma de los 62 términos correspondientes a los 62 cuadros en el tablero de ajedrez truncado es

$$(1 + x)f(x, y) + (1 + y)g(x, y) \qquad \text{(5P.2)}$$

Demuestre que no puede ser posible que (5P. 1) sea igual a (5P.2).

Sugerencia: considere los valores de (5P.1) y (5P.2) para $x = -l$ y $y = -l$.

5.41 Demuestre que independientemente de dónde sean eliminados un cuadro blanco y un cuadro negro de un tablero de ajedrez de 8 x 8, el tablero de ajedrez defectuoso siempre podrá ser cubierto con 31 fichas de dominó de 2 x 1.

5.42 Un tablero de ajedrez de 6 x 6 puede ser embaldosado por 18 fichas de dominó de 2 x 1, como se ilustra en la figura 5P. 17. Se dice que un embaldosado es libre de fallas si cualquier alineamiento vertical (existen 5) y cualquier alineamiento horizontal (existen 5 de ellos) son cruzados al menos por una ficha de dominó (el embaldosado de la figura 5P. 17 no es libre de fallas debido a que el alineamiento horizontal marcado con una flecha no es cruzado por ninguna ficha de dominó). Demuestre que no existe un embaldosado libre de fallas en un tablero de ajedrez de 6 x 6.

Sugerencia: si un alineamiento (vertical u horizontal) es cruzado por fichas de dominó, ¿éste será cruzado por al menos cuántas fichas?

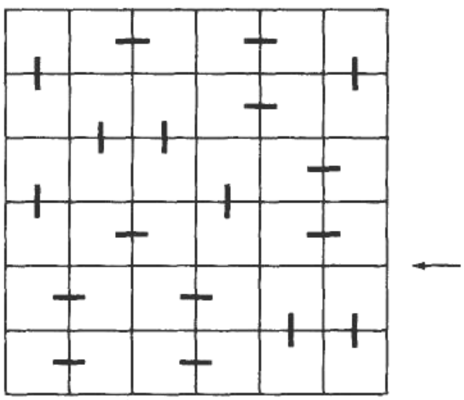


Figura 5P.17

5.43 Un grafo $G = (V, E)$ se llama un grafo *bipartido* si el conjunto de vértices Fpuede particionarse en dos subconjuntos no vacíos X y Y tales que cada arista de E una un vértice de X con un vértice de Y.

a) Sean $G = (V, E)$ un grafo bipartido, X un conjunto de obreros y Y un conjunto de trabajos. Una arista $\{v_x, v_y\}$ de E denota el hecho de que el obrero v_x está calificado para realizar el trabajo v_y. ¿Qué interpretación puede darse a un 1-factor de G?

b) Sean X un conjunto de comités y Y un conjunto de senadores. Una arista $\{v_x, v_y\}$ de E denota el hecho de que el senador v_y es miembro del comité v_x. ¿Qué interpretación puede darse a un 1-factor de G?

c) Sean X un conjunto de niños y 7 un conjunto de niñas. Una arista $\{v_x, v_y\}$ en E denota el hecho de que el niño v_x conoce a la niña v_y. ¿Qué interpretación puede darse a un 1-factor de G?, ¿a un circuito hamiltoniano en G?

5.44 *a)* Demuestre que un grafo aplanable lineal tiene un vértice de grado 5 o menor.

b) Demuestre que un grafo aplanable lineal con menos de 30 aristas tiene un vértice de grado 4 o menor.

5.45 Demuestre que en un grafo aplanable lineal conexo con seis vértices y 12 aristas, cada una de las regiones está acotada por 3 aristas.

5.46 Sea *G* un grafo con 11 o más vértices. Demuestre que *G* o bien *G* no es aplanable.

5.47 El *espesor* de un grafo *G* se define como el número mínimo de subgrafos aplanables (de aristas disjuntas) en los cuales puede ser descompuesto *G*. Denotaremos el espesor de *G* mediante 9(G).

a) ¿Cuál es el espesor de un grafo aplanable?

b) ¿Cuánto es $\theta(K_5)$?, ¿ θ (K_8)?

c) Demuestre que para cualquier grafo lineal sin lazos *G*

$$\theta(G) \geqq \left\lceil \frac{e}{3v-6} \right\rceil$$

d) Demuestre que

$$\theta(K_n) \geqq \left\lfloor \frac{n+7}{6} \right\rfloor$$

Sugerencia: demuestre que $\lceil p/q \rceil = \lfloor (p+q-1)/q \rfloor$ para cualesquiera enteros positivos *p* y *q*.

CAPÍTULO

SEIS

Árboles y conjuntos de corte

6.1 ÁRBOLES

En este capítulo estudiaremos una clase de grafos llamados *árboles.* Consideremos un grupo de boxeadores que pelean por el campeonato de peso completo mundial. Supongamos que cada boxeador tiene una oportunidad para enfrentar al campeón vigente, y que el perdedor de cualquier encuentro será eliminado de la contienda. Sea $G = (V, E)$ un grafo no dirigido donde los vértices de V representan a los boxeadores y las aristas de E representan los encuentros. Esto es, para x y y en V, $\{x, y)$ está en E si hubo un encuentro entre xyy. Por ejemplo, sea $V= \{a, b, c, d, e, f, g, h, i)$. Supongamos que, al inicio a es el campeón vigente. En las confrontaciones siguientes, a venció ab,c y d,y pierde el título frente a e; e venció a f y g y pierde el título frente a h; finalmente, h pierde el título frente a i. El grafo de la figura 6.1a muestra todos los encuentros realizados. Otro ejemplo, consideremos cuatro parejas de chismosos $\{a, A, b, B, c, C, d, D\}$, donde a , b , c y d son los esposos y A, B, C *y* D, son, respectivamente, sus esposas. Supongamos que a llama a su esposa para contarle algún chisme, entonces ella llama a las otras esposas para difundir el chisme, y cada una de ellas, a su vez, llama a su esposo para comunicárselo. El grafo en la figura 6.1b muestra cómo se propagó el chisme, y dónde las aristas representan las llamadas telefónicas. Los grafos en estos ejemplos comparten algunas propiedades comunes, las cuales identificaremos y estudiaremos.

Definimos un *árbol* como un grafo (no dirigido) conexo que no contiene circuitos. Por ejemplo, los grafos en las figuras 6. la y 6.1b son árboles. Una colección de árboles disjuntos es llamado un *bosque.* Un vértice de grado 1 en un árbol se llama una *hoja* o un *nodo terminal*, *y* un vértice de grado mayor que 1 recibe el nombre de un *nodo rama* o *nodo interno.* Por

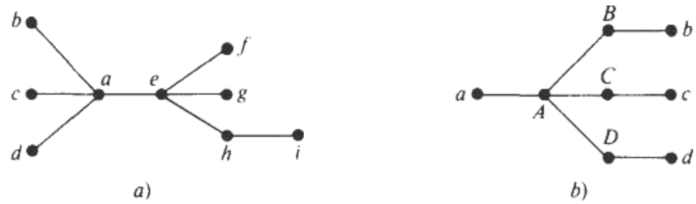


Figura 6.1

ejemplo, los vértices *b, c, d, f, g* e *i en* la figura 6.1a son hojas, y los vértices *a* , *e* y *h* son nodos rama. Señalemos algunas propiedades de los árboles:

1. Existe un único paseo entre dos vértices cualesquiera en un árbol.†
2. El número de vértices es mayor en uno al número de aristas en un árbol.
3. Un árbol con dos o más vértices tiene al menos dos hojas.

La propiedad 1 se deduce de la definición de un árboi. Ya que un árbol es un grafo conexo, existe al menos un paseo entre cualesquiera dos vértices. No obstante, si hubiera dos o más paseos entre un par de vértices, habría un circuito en el árbol. Así hemos probado la propiedad 1. Para los ejemplos en las figuras 6.1 *a y* 6.1 *b* podemos verificar esta propiedad de inmediato.

Demostraremos la propiedad 2 por inducción sobre el número de vértices en el árbol. Como base de la inducción, vemos que un árbol con un vértice no tiene aristas, y un árbol con dos vértices tiene una arista. Consideremos un árbol *T* con v vértices y e aristas.‡ Sea *{a, b}* una arista de *T*. Supongamos que eliminamos la arista *{a, b}* de *T*. Entonces aseguramos que las aristas restantes forman un bosque de dos árboles. Sea c un vértice tal que el paseo entre *a* y *c* en *T* no incluye la arista *{a, b}*. Entonces el paseo entre *b y c* en *T* debe incluir la arista *{a, b}* porque, de otra forma, existe un circuito en *T*. Así, después de eliminar la arista *{a, b}*, hay un paseo entre *a* y *c* pero no un paseo entre *b* y *c*. De modo similar, sea *d* un vértice tal que el paseo entre *a y d* en *T* incluye la arista *{a, b)*. Entonces el paseo entre *b* y *d* no incluye la arista *{a, b}*. En consecuencia, después de eliminar la arista *{a, b}*, existe un paseo entre *b* y *d* pero no un paseo entre *a* y *d*. Por tanto, la eliminación de la arista *{a, b}* divide a *T* en dos árboles disjuntos *T´* y *T''*, donde *T´* contiene a *a* y todos los vértices cuyos paseos hacia *a* en *T* no contienen la arista *{a, b) y T''* contiene a *b* y todos los vértices cuyos paseos hacia *b* en *T* no contienen el arista {*a*, *b)* Como *T'* y *T''* tienen a lo más v - 1 vértices, se sigue de la hipótesis de inducción que

$$\mathbf{e}' = \mathbf{v}' - 1$$
$$\mathbf{e}'' = \mathbf{v}'' - 1$$

† A lo largo de este capítulo usaremos el término *paseo* para dar a entender *un paseo elemental,* a menos que se especifique otra cosa.

‡ Usaremos v para denotar el número de vértices y e para denotar el número de aristas en un grafo.

donde e' y e" son el número de aristas y v" y v" son el número de vértices en *T´* y *T"*. Así,

$$e' + e'' = v' + v'' - 2$$

Puesto que

$$e = e' + e'' + 1$$

$$v = v' + v''$$

tenemos

$$e = v - 1$$

Observemos que el árbol de la figura 6.1*a* tiene nueve vértices y ocho aristas, y el árbol de la figura 6.1*b* tiene ocho vértices y siete aristas.

La propiedad 3 se obtiene de la propiedad 2. Recordemos que la suma de los grados de los vértices en un grafo cualquiera es igual a 2e, lo cual es igual a 2v - 2 en un árbol. Como un árbol con más de un vértice no puede tener ningún vértice aislado, deberá haber al menos dos vértices de grado 1 en el árbol. Ambos ejemplos en la figura 6.1 tienen más de dos hojas. Como ejercicio, invitamos al lector a caracterizar todos los árboles que tienen exactamente dos hojas (véase el problema 6.1).

Ahora demostraremos algunos resultados sobre la caracterización de árboles. Todas estas características pueden considerarse como definiciones equivalentes para árboles:

1. Un grafo en el cual existe un único paseo entre cada par de vértices es un árbol.
2. Un grafo conexo con e = v - 1 es un árbol.
3. Un grafo con e = v - 1 que no tiene circuitos es un árbol.

Para demostrar la caracterización 1, primero señalemos que un grafo en el cual existe un paseo entre cada par de vértices es conexo. Además, el grafo no puede contener un circuito si estos paseos son únicos, puesto que la existencia de un circuito implica la existencia de dos paseos distintos entre un cierto par de vértices. Así, podemos concluir que un grafo en el cual existe un paseo único entre cada par de vértices es un árbol.

Como ejemplo consideremos a los espías en una red de espionaje, organizada de tal manera que cada dos espías pueden comunicarse uno con otro ya sea directamente o a través de una cadena única de sus colegas. Sea *V* el conjunto de espías y £ el conjunto de aristas tal que si la arista *{a, b}* está en *E* significa que los espías *a* y *b* pueden comunicarse uno con otro directamente. De acuerdo con la caracterización 1, *G = (V, E)* es un árbol.

Ahora demostraremos la caracterización 2. Sea *G* un grafo conexo con e = v - 1. Supongamos que *G* contiene un circuito simple *C*.† Denotemos el número de vértices de *C* por *c*. Es obvio que el número de aristas de *C* es igual a *c*. Dado que *G* es conexo, cada vértice de *G* que no está en *C* debe estar conectado a los vértices de *C*. Ahora cada arista de *G* que no está en C puede conectar sólo otro vértice adicional a los vértices de *C*. Existen v - c

† Si *G* contiene un circuito, *G* contiene un circuito simple.

vértices que no están en *C*, de manera que *G* debe contener al menos v - c aristas que no están en *C*. Así, debemos tener que $\mathbf{e} \geqq \mathbf{c} + (\mathbf{v} - \mathbf{c}) = \mathbf{v}$, lo cual es una contradicción. De esto se concluye que *G* no contiene ningún circuito, y por tanto es un árbol.

Consideremos un ejemplo en el cual alguien envía un lote de cartas a sus amigos. Sus amigos, a su vez, envían cartas a sus amigos quienes no han recibido alguna carta, y así sucesivamente. Sea *G = {V, E)* un grafo donde *V* es el conjunto de personas que han recibido una carta junto al iniciador, y *E* es el conjunto de aristas que representan las cartas. Debido a que cada carta agrega una persona al grupo de quienes han recibido cartas, tenemos que e = v - 1. Además, *G* es un grafo conexo debido a que sólo hay un iniciador de las cartas. Así, de acuerdo con la caracterización 2, *G* es un árbol.

Enseguida demostraremos la caracterización 3. Sea *G* un grafo con e = v - 1 que no tiene un circuito. Supongamos que *G* no es conexo. Denotemos por *G', G''*, . . . a las componentes conexas de *G*. Como cada una de las *G', G ''*,... es conexa y no tiene circuitos, todas ellas son árboles. De acuerdo con la propiedad 2 de los árboles, demostrada con anterioridad, tenemos

$$\mathbf{e}' = \mathbf{v}' - 1$$

$$\mathbf{e}'' = \mathbf{v}'' - 1$$

$$\cdots$$

donde e', e'', . . . son el número de aristas y v', v'', . . . son el número de vértices en *G'*, *G ''*, ... Tenemos que

$$e' + e'' + ... = v' - 1 + v'' - 1 + \dots \tag{6.1}$$

Puesto que

$$e = e' + e'' + \dots$$

$$y = y' + y'' + \dots$$

(6.1) implica que

$$e < v - 1$$

lo cual es una contradicción. Así, *G* debe ser conexo, y por tanto es un árbol.

Consideremos el ejemplo presentado al inicio, sobre el grupo de boxeadores que pelean por el campeonato de peso completo. Queremos demostrar que el grafo que describe los encuentros, G, siempre es un árbol para cualquier grupo de boxeadores y para cualesquiera posibles resultados de los encuentros. Puesto que cada encuentro elimina exactamente a uno de los boxeadores de la contienda, tenemos e = v - 1. Supongamos que tenemos un circuito en *G*. Denotemos por (c_1 c_2, c_3,..., c_k) la sucesión de vértices en el circuito. *Sin pérdida de generalidad*, podemos suponer que c_2 es el perdedor del encuentro entre c_1 y c_2. En ese caso, c_2 debe ser el vencedor y c_3 el perdedor en un encuentro anterior. De modo similar, c_3 debe ser el vencedor y c_4 debe ser el perdedor en un encuentro anterior al encuentro entre c_1 y c_2. Por último, c_k debe ser el vencedor y c_x debe ser el perdedor en un encuentro anterior al encuentro entre c_x y c_2, lo cual es imposible. Así, de acuerdo con la caracterización 3, *G* es un árbol.

6.2 ÁRBOLES ENRAIZADOS

Diremos que un grafo dirigido es un *árbol dirigido* si se convierte en un árbol cuando se ignoran las direcciones de sus aristas. Por ejemplo, la figura 6.2*a* muestra un árbol dirigido. Un árbol dirigido es un *árbol enraizado* si existe exactamente un vértice cuyo grado de entrada sea 0 y los grados de entrada de todos los otros vértices sean 1. El vértice con el grado de entrada 0 es llamado la *raíz* del árbol enraizado. Por ejemplo, la figura 6.2*b* muestra un árbol enraizado. En un árbol enraizado, un vértice cuyo grado de salida sea 0 se llama una *hoja o* un *nodo terminal,* y un vértice cuyo grado de salida sea diferente de cero se llama un *nodo rama* o un *nodo interno.* En muchas ocasiones encontramos estructuras que pueden representarse como árboles enraizados. Por ejemplo, el organigrama de una corporación como el de la figura 6.3 puede representarse inmediatamente por un árbol enraizado.

Sea *a* un nodo rama en un árbol enraizado. Diremos que un vértice *b* es un *hijó^* de *a* si existe una arista de *a* a *b.* Además, se dice que *a* es *elpadre* de *b.* Dos vértices son *hermanos* si son hijos del mismo vértice. Diremos que un vértice *c* es un *descendiente* de *a* si existe un paseo dirigido de *a* a *c.* Además, se dice que *a* es un *ancestro* de *c*. Estos términos nos recuerdan lo que llamamos árboles familiares que en realidad son árboles enraizados.

Sea *a* un nodo rama en un árbol *T.* Por *el subárbol con raíz a* entendemos el subgrafo *T´* = (*V, E*') de *T* tal que *V´* contiene a *a* y a todos sus descendientes y *E'* contiene las aristas de todos los paseos dirigidos que surgen de *a.* Por un *subárbol de a* entendemos un subárbol que tiene a un hijo de a como raíz.

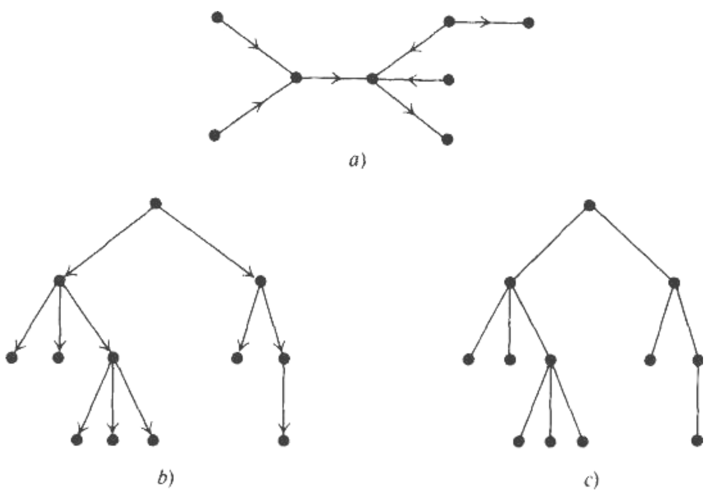


Figura 6.2

† En la literatura, los términos *hija y niño* también son utilizados. No nos preocuparemos sobre la discriminación por el sexo.

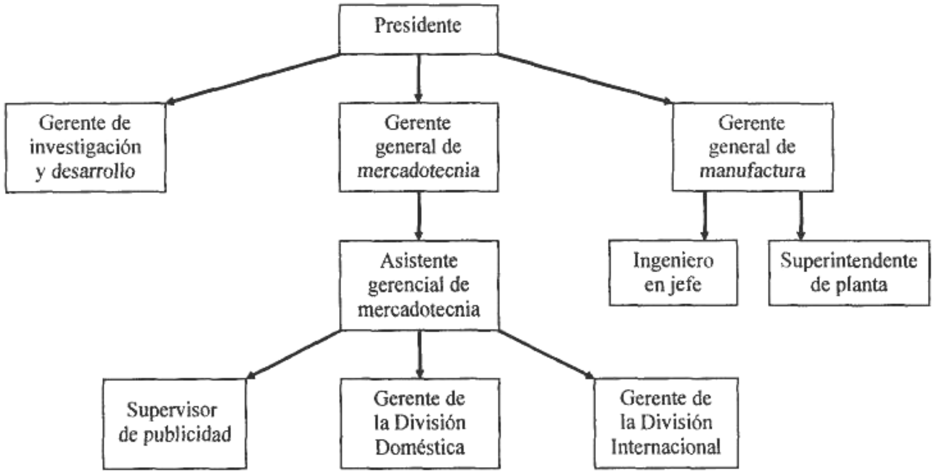


Figura 6.3

Cuando trazamos un árbol enraizado, si nos apegamos a la convención de colocar los hijos de un nodo rama bajo éste, las puntas de flecha de las aristas pueden omitirse, debido a que puede entenderse que apuntan hacia abajo. La figura 6.2*c* muestra un ejemplo.

Consideremos el árbol enraizado de la figura 6.4*a* el cual es el árbol familiar de un hombre que tiene dos hijos, de los cuales el mayor no tiene hijos y el menor tiene tres hijos. A pesar de que el árbol enraizado de la figura 6.4*b* es isomorfo al de la figura 6.4*a*, podría

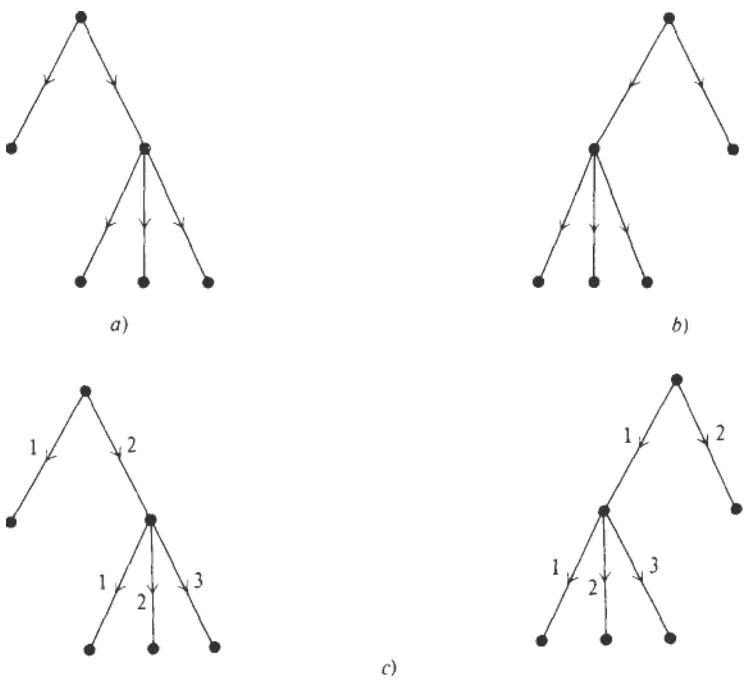


Figura 6.4

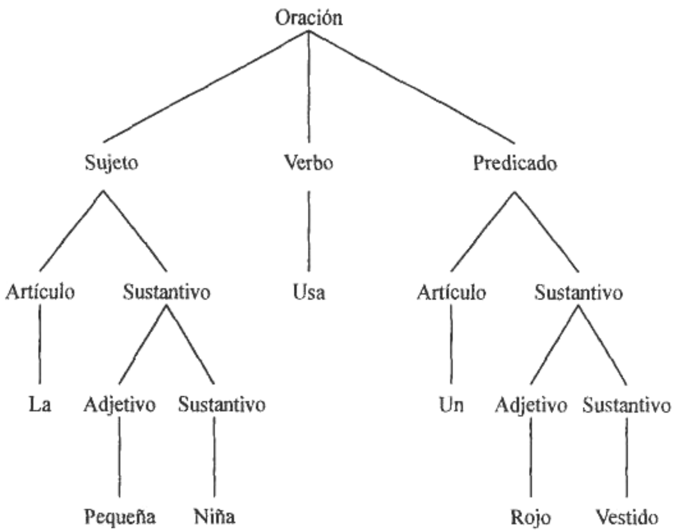


Figura 6.5

ser el árbol familiar de otro hombre cuyo hijo mayor tiene tres hijos y el hijo menor no tiene hijos. Este ejemplo motiva la definición de un árbol ordenado, lo cual nos permitirá referirnos sin ambigüedades a cada uno de los subárboles de un nodo rama. Un *árbol ordenado* es un árbol enraizado con las aristas incidentes desde cada nodo rama, etiquetadas con los enteros $1, 2, \ldots,$ *i*, ... † Por tanto, los subárboles de un nodo rama pueden ser referidos como el primero, el segundo, . . . , y el *i*-ésimo subárboles del nodo rama, y corresponden a las etiquetas de las aristas incidentes desde el nodo (nótese que no insistimos en que las etiquetas sean enteros consecutivos. Por ejemplo, si las tres aristas incidentes desde un nodo rama son etiquetadas como 1,2,5, diremos que el nodo no tiene un tercer ni un cuarto subárbol). Ahora supongamos que los árboles enraizados de las figuras 6.4*a* y 6.4*b* son etiquetados como se muestra en la figura 6.4*c*. Diremos que dos árboles ordenados son isomorfos si existe una correspondencia uno a uno entre sus vértices y aristas, la cual conserva la relación de incidencia, y si las etiquetas de las aristas correspondientes coinciden. Así, los dos árboles ordenados de la figura 6.4*c* no son isomorfos. Un árbol ordenado en el cual cada nodo rama tiene a lo más *m* hijos es llamado un *árbol m-ario*‡ Diremos que un árbol *m*-ario es *regular* si cada uno de sus nodos rama tiene exactamente *m* hijos. Una clase importante de árboles *m*-arios son los *árboles binarios.* En los árboles binarios en lugar de referirnos al primer o al segundo subárbol de un nodo rama, a menudo nos referimos al *subárbol izquierdo* o al *subárbol derecho* del nodo.

---

† Dichas etiquetas pueden estar implícitas cuando sean obvias a partir de la forma en que el árbol fue definido o trazado.

‡ Una notación correspondiente puede ser definida para todos los árboles enraizados.

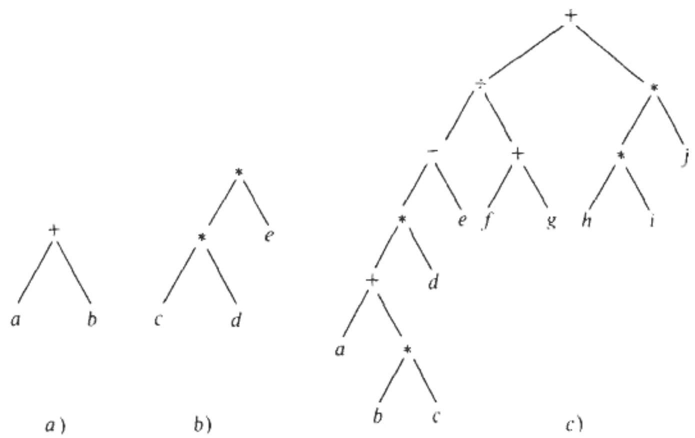


Figura 6.6

El lector ha observado, probablemente, cómo la estructura de una proposición puede representarse por un árbol ordenado. La figura 6.5 muestra un ejemplo según su estructura gramatical en lengua inglesa. Por otra parte, observemos cómo una expresión aritmética puede representarse por un árbol binario. Es claro que expresiones simples como $a + b$ y $c * d * e$ pueden representarse como se muestra en las figuras 6.6*a* y 6.6*b*, donde los operandos aparecen en las hojas y los operadores en los nodos rama de los árboles ordenados. Señalemos que los paréntesis no son necesarios cuando una expresión aritmética es representada por un árbol binario. Por ejemplo, la expresión aritmética $(((a + b * c) * d - e)/(f + g)) + h* i *]$ se representa como se muestra en la figura 6.6*c*.

6.3 LONGITUDES DE PASEOS EN ÁRBOLES ENRAIZADOS

Consideremos el problema de determinar el número de encuentros jugados en un torneo de tenis de eliminación simple.[†] Supongamos que hay ocho competidores en el torneo. Se jugarán cuatro encuentros en la primera ronda, dos encuentros en la segunda ronda, y el encuentro final, lo que da un total de siete encuentros. El problema parece ser más complicado cuando se tiene un número impar de competidores, digamos 11, en el torneo. En este caso, habrá cinco encuentros en la primera ronda (un competidor pasa automáticamente a la siguiente ronda), tres encuentros en la segunda ronda, un encuentro en la tercera (uno de los tres competidores restantes después de la segunda ronda pasa automáticamente al encuentro final), y un encuentro en la ronda final para un total de 10 encuentros. Existe, no obstante, una forma más directa para determinar el resultado. Si examinamos la programación de un torneo de eliminación simple como el que se muestra en la figura 6.7, de inmediato nos damos cuenta de que es exactamente un árbol binario regular en el cual las hojas representan a los competidores en el torneo y los nodos rama representan a los ganadores de los

† En un torneo de eliminación simple un competidor será eliminado después de una derrota.

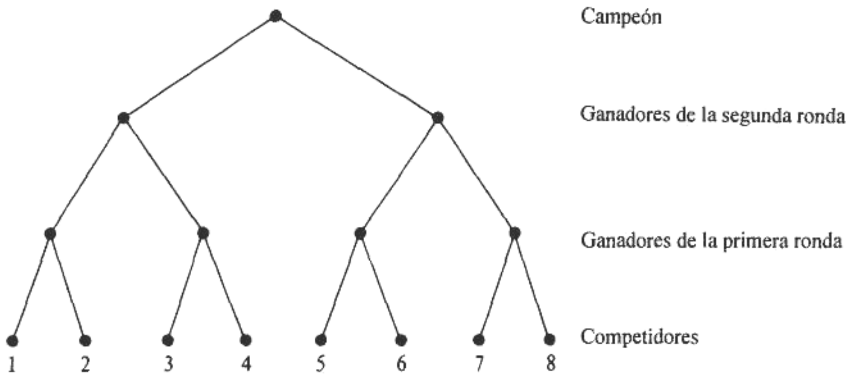


Figura 6.7

encuentros o, equivalentemente, los encuentros jugados en el torneo. Recíprocamente, observamos que cualquier árbol binario regular puede ser visto como la programación de un torneo de eliminación simple.† Así, nos gustaría conocer la relación entre i, el número de nodos rama, y t, el número de hojas, de un árbol binario regular. Puesto que en un torneo de eliminación simple en cada encuentro jugado se elimina un competidor, y al final del torneo, todos los competidores excepto el campeón han sido eliminados, el número de encuentros jugados es uno menos que el total de competidores en el torneo. Por tanto, tenemos

$$i = t - 1$$

El resultado puede extenderse al caso de árboles m-arios regulares. Imaginemos un cierto tipo de encuentro con /«-competidores que tienen un solo ganador. En tal caso, cualquier árbol w-ario regular puede verse como una programación de un torneo de eliminación simple. De nuevo, ya que cada encuentro jugado elimina m - 1 competidores y sólo el campeón queda al final del torneo, tenemos*

$$(m-1)i = t - 1 \tag{6.2}$$

Consideremos algunos ejemplos:

Ejemplo 6.1

El problema es conectar 19 lámparas a un solo contacto eléctrico mediante el uso de cables de extensión, cada uno de los cuales tiene cuatro contactos. Ya que cualquiera de esas conexiones es un árbol cuaternario con el contacto simple conectado a la raíz del árbol, de acuerdo con (6.2),

$$(4-1)i = 19 - 1$$

Esto es, aunque existen muchas maneras de conectar las lámparas, siempre se necesitarán seis cables de extensión.

❐

† Por un torneo de eliminación simple, entendemos *solamente* un torneo en el cual un competidor será eliminado después de una derrota. No insistimos en que cada competidor juegue en la primera ronda, por ejemplo.

‡ Como prueba alternativa, *mi* es el número total de hijos de los nodos rama, lo cual es igual al número de nodos rama y nodos terminales menos uno (la raíz).

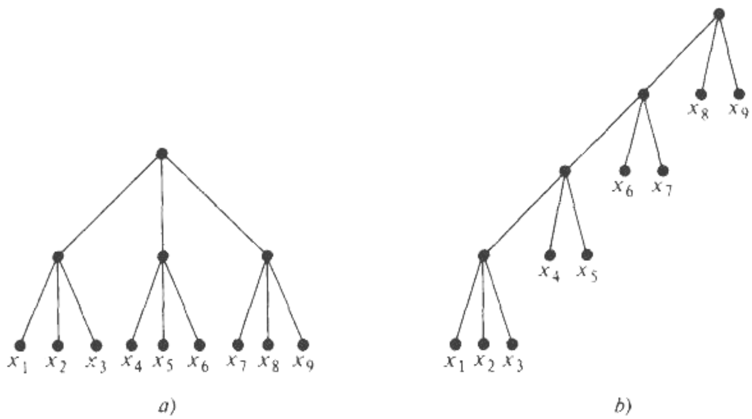


Figura 6.8

Ejemplo 6.2

Consideremos una computadora hipotética que tiene una instrucción que calcula la suma de tres números. Supongamos que queremos encontrar la suma de nueve números, x_h x_2, . . . , x_9. Nos damos cuenta de que cualquier sucesión que permita ejecutar esta instrucción para obtener el resultado es un árbol ternario regular con nueve hojas. Las figuras 6.8*a* y 6.8*b* muestran dos posibles sucesiones. De acuerdo con (6.2),

$$(3 - l)\ i = 9 - 1$$

Por lo que se obtiene

$$i = 4$$

Esto es, la instrucción de suma siempre se ejecutará cuatro veces. ❐

Existe otro aspecto del problema en el ejemplo 6.2 que deseamos analizar. Es claro que dos árboles ternarios diferentes con nueve hojas corresponden a diferentes órdenes en los cuales los nueve números se suman, como se muestra en las figuras 6.8*a* y 6.8*b*. A pesar que siempre se da el caso de que se necesitan al menos cuatro operaciones de adición para calcular la suma $x_1 + x_2 + . . . + x_9$, la posibilidad de calcular algunas sumas parciales simultáneamente (como en el caso de un sistema de cómputo de multiprocesadores), hace que una sucesión sea superior a otra. Por ejemplo, la sucesión de adiciones de la figura 6.8*a* puede llevarse a cabo en dos pasos, donde en cada paso calculamos todas las sumas parciales que pueden efectuarse en forma simultánea. Por otro lado, toma cuatro pasos llevar a cabo la sucesión de adiciones mostrada en la figura 6.8*b*. La *longitud de un paseo* para un vértice en un árbol enraizado se define como el número de aristas en el paseo desde la raíz hasta el vértice. Por ejemplo, la longitud del paseo del vértice x_1 en la figura 6.8*b* es 4, y la del vértice x_5 es 3. Definimos la *altura* de un árbol como el máximo de las longitudes de los paseos en el árbol. Es claro que un árbol /w-ario de altura h tiene a lo más m^h hojas, que corresponde a la figura 6.8a. Por otro lado, un árbol m-ario regular de altura h tiene al menos $m + (m - 1)(h - 1)$ hojas, caso que se muestra en la figura 6.8*b*.

Sean I la suma de las longitudes de los paseos de todos los nodos rama y E la suma de las longitudes de los paseos de las hojas en un árbol enraizado. Demostraremos que para un árbol binario regular $E = I + 2i$, donde i es el número de nodos rama. Consideremos una arista (x, y) en un árbol binario regular T. La arista (x, y) está incluida en el cálculo de las longitudes de los paseos de todos los nodos rama y hojas de T en el subárbol que tiene a y como raíz. Puesto que existe exactamente una hoja más que nodos rama en el subárbol, esta arista se encuentra una vez más en el cálculo de E que en el cálculo de I. Si repetimos este argumento para todas las aristas en el árbol binario, obtenemos

$$\begin{aligned} E &= I + \text{número de aristas en el árbol binario} \\ &= I + 2i \end{aligned}$$

Dejamos al lector llevar a cabo la extensión obvia para demostrar que, para un árbol m-ario regular,

$$E = (m - 1)I + mi$$

6.4 CÓDIGOS DE PREFIJOS

Ahora presentaremos más ejemplos sobre el concepto de longitudes de paseos en árboles enraizados. Consideremos un problema de telecomunicaciones en el cual queremos representar las letras del alfabeto inglés mediante sucesiones de números 0 y 1 (o mediante sucesiones de líneas y puntos). Debido a que las 26 letras del alfabeto deben ser representadas por sucesiones distintas de números 0 y 1, se pueden representar por sucesiones de longitud 5 ($2^4 < 26 < 2^5$). Para enviar un mensaje, simplemente transmitimos una cadena de números 0 y 1 que contenga las sucesiones para las letras en el mensaje. En el extremo receptor, esta cadena de números 0 y 1 será dividida en una sucesión de longitud 5 y las letras correspondientes podrán ser reconocidas.

Es bien conocido, no obstante, que las letras del alfabeto no se usan con frecuencias uniformes. Por ejemplo, las letras *e* y *t* se usan más frecuentemente que las letras *q* y *z*. Así, uno podría desear representar las letras más frecuentemente usadas con sucesiones más cortas, y las letras menos frecuentemente usadas con sucesiones más largas, de manera que se reduzca la longitud total de la cadena. Entonces surge un problema interesante: cuando representamos las letras por sucesiones de varias longitudes, existe la pregunta de cómo ha de dividirse en el extremo receptor, sin ambigüedades, una cadena larga de números 0 y 1 en sucesiones que correspondan a las letras. Por ejemplo, si usamos las sucesiones 00 para representar la letra e, 01 para representar la letra *t*, y 0001 para representar la letra *w*, no estaremos en condiciones de determinar si el texto transmitido fue *et* o *w* cuando recibimos la cadena 0001 en el extremo receptor. Diremos que un conjunto de sucesiones es un *código de prefijos* si no existe una sucesión en el conjunto que sea un prefijo de otra sucesión del conjunto. Por ejemplo, el conjunto {000,001,01,10,11} es un código de prefijos, en tanto que el conjunto {1,00,01,000,0001} no lo es. Mostraremos que si representamos las letras del alfabeto por sucesiones en un código de prefijos, siempre será posible dividir la cadena recibida en sucesiones que representen las letras en un mensaje sin ambigüedades.

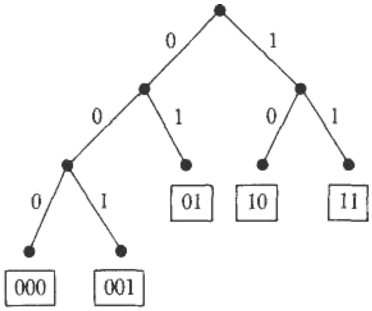


Figura 6.9

Señalemos, primero, que podemos obtener un código de prefijos binario directamente a partir de un árbol binario. Para un árbol binario dado, etiquetamos las dos aristas incidentes desde cada nodo rama con 0 y 1. Asignamos a cada hoja una sucesión de números 0 y 1, que es la sucesión de etiquetas para las aristas del paseo desde la raíz hasta la hoja. Por ejemplo, la figura 6.9 muestra un árbol binario y las sucesiones asignadas a sus hojas, las cuales encerramos por claridad en rectángulos. Es evidente que el conjunto de sucesiones asignadas a las hojas en cualquier árbol binario es un código de prefijos.

Recíprocamente, queremos demostrar que existe un árbol binario que corresponde a un código de prefijos, con las dos aristas incidentes desde cada uno de los nodos rama etiquetados con 0 y 1, de manera que las sucesiones de números 0 y 1 asignadas a las hojas sean las sucesiones en el código.* Sea h la longitud de la(s) sucesión(es) más larga(s) en el código de prefijos. Construimos un árbol binario regular *lleno* de altura h[‡] y etiquetamos las dos aristas incidentes desde cada uno de los nodos rama con 0 y 1. Asignamos a cada vértice una sucesión de números 0 y 1 que es la sucesión de etiquetas de las aristas en el paseo desde la raíz hasta el vértice. Es claro que a cada sucesión binaria de longitud menor o igual a h es asignada exactamente a un vértice. Marquemos todos los vértices que han sido asignados con sucesiones en el código de prefijos, y entonces podamos el árbol para eliminar cada vértice, con la arista incidente a éste, que no tenga un descendiente marcado. Puesto que ningún vértice marcado es un antepasado de otro vértice marcado, el conjunto de vértices marcados será exactamente el conjunto de hojas del árbol resultante. Éste es, en efecto, el árbol que buscábamos. Como ejemplo, para el código de prefijos {1,01, 000, 001} obtenemos el árbol binario de la figura 6.10*b* al podar el árbol binario regular lleno de altura 3 de la figura 6.10*a*, donde los vértices que son asignados a las sucesiones en el código de prefijos están encerrados en rectángulos.

Por último, estamos listos para dar un argumento simple que demuestre que siempre es posible dividir una cadena de números 0 y 1 recibida en sucesiones que están en un código de prefijos. Si comenzamos en la raíz del árbol binario, trazamos hacia abajo un paseo en el árbol de acuerdo con los dígitos de la cadena recibida. Esto es, en un nodo rama seguiremos

† Para una prueba alternativa, véase el problema 6.11.

‡ Un árbol binario regular *lleno* es un árbol regular en el cual todas las hojas tienen la misma longitud de paseo (igual a la altura del árbol).

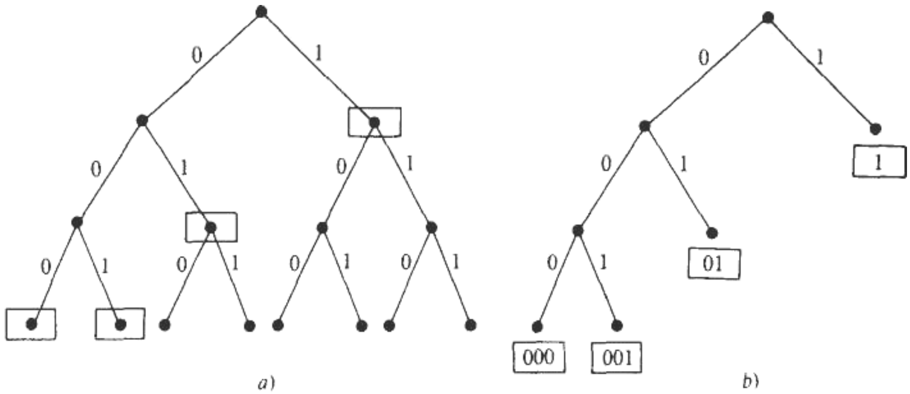


Figura 6.10

la arista etiquetada como un 0 si encontramos un 0 en la cadena recibida, y seguiremos la arista etiquetada con 1 si encontramos un 1 en la cadena recibida. Cuando el paseo hacia abajo alcanza una hoja, sabemos que ha sido detectada una sucesión en el código de prefijos, y deberemos regresar a la raíz del árbol para comenzar a buscar la siguiente sucesión. Este procedimiento garantiza que no habrá ambigüedad al dividir la cadena recibida en sucesiones que están en el código de prefijos.

Otro problema interesante es ser más preciso acerca de la idea de usar sucesiones cortas para representar las letras usadas con más frecuencia. Para esto necesitamos alguna información sobre las frecuencias del uso de las letras. Desde luego, esa información está disponible, y la tabla 6.1 muestra el número promedio de apariciones de las letras en el alfabeto inglés.[1'] Si usamos una sucesión binaria de /, dígitos para representar la /-ésima letra del

Tabla 6.1

Letra	Número de apariciones en 1 000 letras	Letra	Número de apariciones en 1 000 letras
a	82	n	71
b	14	o	80
c	28	p	20
d	38	q	1
e	131	r	68
f	29	s	61
g	20	t	105
h	53	u	25
i	63	v	9
j	1	w	15
k	4	x	2
l	34	y	20
m	25	z	1

† Véase, por ejemplo, Zipf [14].

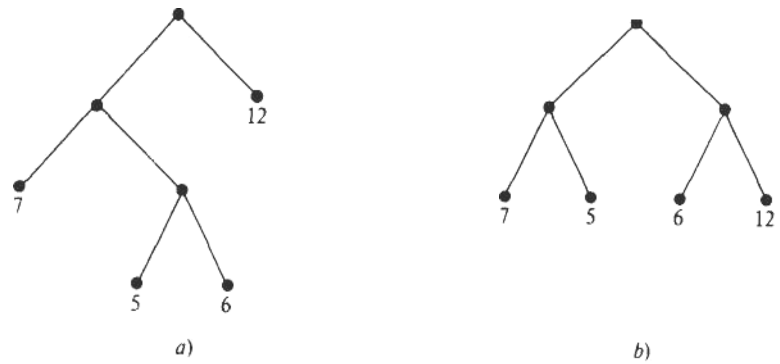


Figura 6.11

alfabeto, la longitud promedio de la cadena binaria que representa un texto en inglés de 1 000 letras será $\Sigma_{i=1}^{26} w_i l_i$, donde w, es el número promedio de apariciones de la *i*-ésima letra en las 1 000 letras.

Nuestro análisis nos conduce al siguiente problema: supongamos un conjunto de pesos $w_1, w_2, \ldots, w_t$. Sin pérdida de generalidad, supongamos que $w_1 \leqq w_2 \leqq \cdots \leqq w_t$. Diremos que un árbol binario que tiene *t* hojas con los pesos $w_1, w_2, \ldots, w_t$ asignados a las hojas es un *árbol binario para los pesos* $w_1, w_2, \ldots, w_t$. Definimos el *peso de un árbol binario para los pesos* $w_1, w_2, \ldots, w_t$ como es $\Sigma_{i=1}^{t} w_i \, l(w_i)$, donde $l(w_i)$ la longitud del paseo de la hoja a la cual se le asignó el peso w_i. El peso del árbol *T* se denotará por *W(T).* Diremos que un árbol binario para los pesos $w_1, w_2, \ldots, w_t$ es un árbol óptimo si su peso es mínimo. Por ejemplo, dados los pesos 5,6, 7 y 12, la figura 6.11*a* muestra un árbol óptimo (exhortamos al lector a verificar si el árbol de la figura 6.11*b* es óptimo).

Existe un procedimiento muy elegante debido a D. A. Huffman [6] para construir un árbol óptimo para un conjunto dado de pesos. La observación clave que conduce al procedimiento de construcción, es que podemos obtener un árbol óptimo para los pesos w_1, $w_2, \ldots, w_t$, a partir de un árbol óptimo *T'* para los pesos $w_1 + w_2, w_3, w_4, \ldots, w_t$.† Específicamente, aseguramos que al remplazar la hoja en 7'', a la cual se le asigna el peso $w_1 + w_2$ por el subárbol mostrado en la figura 6.12, originaremos un árbol óptimo para los pesos $w_1, w_2, \ldots, w_t$. Por ejemplo, supongamos que queremos construir un árbol óptimo para los pesos.3, 4, 5, 6 y 12. Como sabemos que el árbol de la figura 6.11*a* es un árbol óptimo para los pesos 5,6,7 y 12, obtenemos el árbol de la figura 6.13 como un árbol óptimo para los pesos 3,4,5,6 y 12. Para demostrar nuestra afirmación, primero queremos demostrar

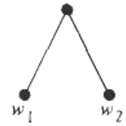


Figura 6.12

† Observemos que los pesos $w_1 + w_2, w_3, w_4, \ldots, w_t$ podrían no encontrarse en un orden creciente.

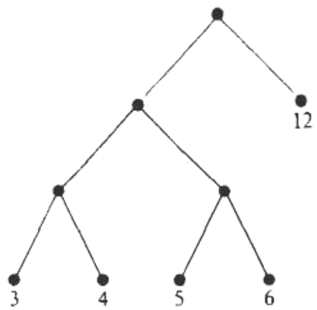


Figura 6.13

que existe un árbol óptimo para los pesos $w_1, w_2, \ldots, w_t$ en el que las hojas a las cuales son asignados w_1 y w_2 son hermanas. Supongamos que tenemos un árbol óptimo para los pesos $w_1, w_2, \ldots, w_t$. Sea *a* el nodo rama con la longitud del paseo más largo en el árbol. Supongamos que los pesos asignados a los hijos de *a* son w_x y w_y. Así, $l(w_x) \geqq l(w_1)$ y $l(w_x) \geqq l(w_2)$. Por otro lado, como el árbol es óptimo, debemos tener $l(w_x) \leqq l(w_1)$ y $l(w_x) \leqq l(w_2)$. [Si $l(w_x) > l(w_1)$, al intercambiar las asignaciones de w_1 y w_x damos lugar a un árbol de peso similar, lo cual es una contradicción. El mismo argumento se aplica si $l(w_x) > l(w_2)$.] Por tanto, tenemos que $l(w_x) = l(w_y) = l(w_1) = l(w_2)$. Si intercambiamos las asignaciones de w_x, w_y con las de w_1, w_2 obtenemos un árbol óptimo en el que las hojas a las cuales w_1 y w_2 han sido asignadas son hermanas.

Denotemos por $\hat{T}$ a un árbol óptimo para los pesos $w_1, w_2, \ldots, w_t$ en el cual las hojas asignadas a w_1 y w_2 son hermanas. Remplacemos el subárbol de $\hat{T}$ que contiene estas dos hojas y su padre por una hoja, y asignemos a esta nueva hoja el peso $w_1 + w_2$. Denotemos por $\hat{T}'$ este árbol resultante, el cual es un árbol binario para los pesos $w_1, w_2, w_3, \ldots, w_t$. Es evidente que

$$W(\hat{T}) = W(\hat{T}') + w_1 + w_2$$

Sea T' un árbol óptimo para los pesos $w_1, w_2, w_3, \ldots, w_t$. Sea T el árbol obtenido a partir de T' mediante el remplazo de la hoja de T' a la cual se le asignó $w_1 + w_2$ por el subárbol mostrado en la figura 6.12. Tenemos

$$W(T) = W(T') + w_1 + w_2$$

Si $W(T) > W(\hat{T})$, entonces $W(T') > W(\hat{T}')$, lo cual es una contradicción porque T' es un árbol óptimo para los pesos $w_1 + w_2, w_3, \ldots, w_t$. Por tanto, T es un árbol óptimo para los pesos $w_1, w_2, \ldots, w_t$.

De acuerdo con nuestra observación, el problema de construir un árbol óptimo para *t* pesos se puede reducir al de construir uno para *t* -1 pesos, el cual puede reducirse al construir uno para *t* - *2* pesos, y así sucesivamente. Puesto que el problema de construir un árbol óptimo para dos pesos es trivial, el problema de construir un árbol óptimo para *t* pesos está resuelto. Por ejemplo, para los pesos 3,4, 5, 6 y 12, la figura 6.14 muestra la construcción paso por paso de acuerdo con el procedimiento de Huffman.

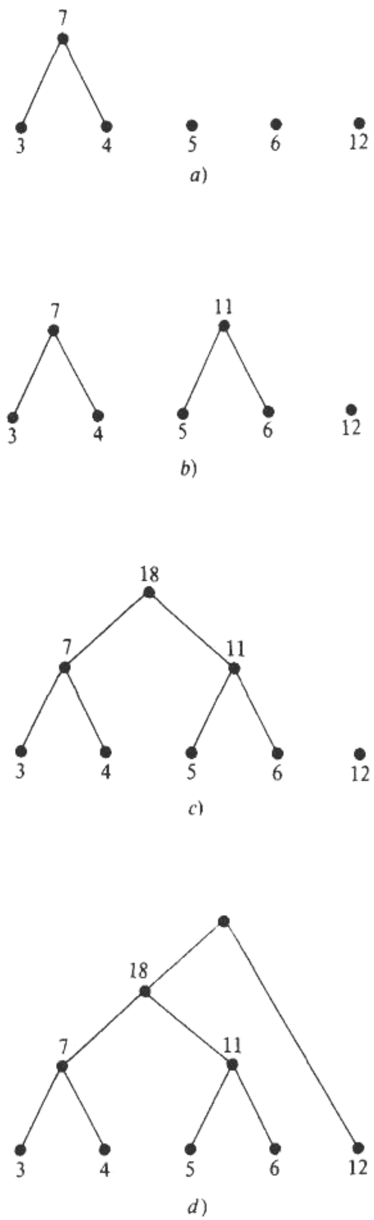


Figura 6.14

6.5 ÁRBOLES DE BÚSQUEDA BINARIA

Supongamos que un amigo nuestro ha seleccionado un número entre 1 y 99, y queremos determinar de qué número se trata con preguntas del siguiente estilo, "¿el número es 37?". Nuestro amigo responderá indicándonos si efectivamente es 37, o el número es menor que 37, o el número es mayor que 37. Aunque podemos determinar el número que nuestro amigo ha seleccionado por medio de la serie de preguntas, "¿el número es 1?, ¿el número es 2?, ¿el número es 3?,..., ¿el número es 99?", la mayoría de nosotros emplearía la siguiente estrategia: comenzamos con la pregunta, "¿el número es 50?". Si se nos indica que el número es 50, hemos terminado. Si se nos indica que el número es menor que 50, hemos reducido el intervalo que contiene al número del 1 a 49. Si se nos indica que el número es mayor que 50, hemos reducido el intervalo que contiene al número del 51 a 99. De acuerdo con esto, si el número se encuentra entre 1 y 49, nuestra siguiente pregunta es, "¿el número es 25?"; y si el número se encuentra entre 51 y 99, nuestra pregunta será "¿el número es 75?". Es bastante obvio que nuestra estrategia es realizar una serie de preguntas de manera que la respuesta de nuestro amigo a cada pregunta nos permita determinar el número que ha seleccionado, o reducir el intervalo que contiene al número a la mitad de lo que teníamos anteriormente. Confiamos en que el lector estará de acuerdo con que ésta es una buena estrategia.

Antes de continuar presentamos otros dos ejemplos. Supongamos que los registros de los empleados de una compañía se acomodan de acuerdo con los números de seguridad social. Un problema interesante consiste en diseñar un procedimiento para la búsqueda del registro de un empleado, dado su número de seguridad social. Un problema similar es encontrar el número telefónico de una persona en un directorio telefónico o determinar que el número no está en el directorio. Es claro que la estrategia en el párrafo precedente puede aplicarse inmediatamente a estos casos.

Formulemos con precisión el problema de buscar un objeto en una lista ordenada. Suponemos que trabajamos con objetos sobre los cuales existe un ordenamiento lineal, $<$. En ejemplos prácticos, el ordenamiento lineal puede ser numérico, alfabético, alfanumérico, etcétera. Sean $K_1, K_2, \ldots, K_n$ los n objetos de una lista ordenada los cuales son conocidos como las *claves*. Supongamos que $K_1 < K_2 < \ldots < K_n$. Dado un objeto x, nuestro problema es buscar las claves y determinar si x es igual a una de las claves o si x está entre las claves K_i y K_{i+1} para algún i. Primero señalemos que la búsqueda tiene $2n + 1$ posibles resultados, es decir, x es menor que K_1, x es igual a K_1, es mayor que K_1 pero menor que K_2, x es igual a K_2, etcétera.

Un procedimiento de búsqueda consiste en una serie de comparaciones entre x y las claves donde cada comparación de x con una clave nos indica si x es igual, menor que, o mayor que tal clave.† Ahora mostraremos cómo podemos describir un procedimiento de búsqueda mediante una representación de árbol binario. Definimos *un árbol de búsqueda* para las claves $K_1, K_2, \ldots, K_n$ como un árbol binario con n nodos rama y $n + 1$ hojas. Los nodos rama son etiquetados con $K_1, K_2, \ldots, K_n$, y las hojas son etiquetadas con $K_0, K_1, K_2, \ldots, K_n$‡ de manera que para el nodo rama con la etiqueta K_i, su subárbol izquierdo contiene sólo los vértices con etiquetas K_j, $j < i$, y su subárbol derecho contiene sólo vértices con etiquetas K_j, $j \geq i$. Por ejemplo, la figura 6.15 muestra un árbol de búsqueda para las claves K_1, K_2, K_3, K_4, donde, por claridad, usamos círculos para denotar nodos rama y cuadrados para denotar hojas. De inmediato se ve que un árbol de búsqueda corresponde a un proce-

† Supongamos que comparamos x con K_i y encontramos que x es menor que K_i. Un paso siguiente para comparar x con K_j para $j > i$ sería un desperdicio total. Suponemos que los procedimientos de búsqueda considerados no contienen ninguno de esos pasos desperdiciados.

‡ El significado de la etiqueta A, será aclarado mas adelante.

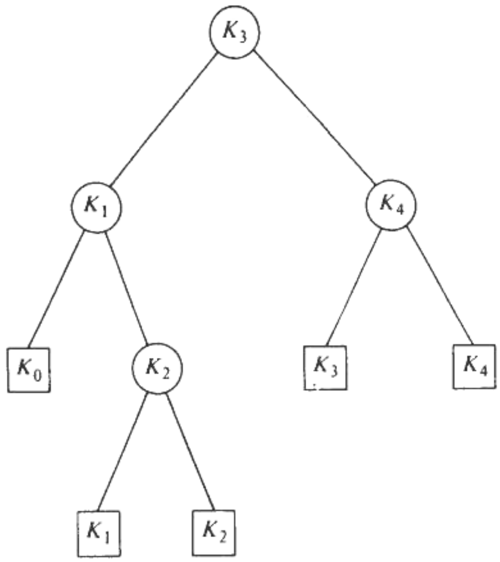


Figura 6.15

dimiento de búsqueda: al comenzar con la raíz del árbol de búsqueda, comparamos un obj eto dado x con la etiqueta de la raíz K_i. Si x es igual a K_i, la búsqueda ha terminado. Si x es menor que K_i, comparamos x con el hijo izquierdo de la raíz, y si x es mayor que K_i, comparamos x con el hijo derecho de la raíz.† Dicha comparación se continúa para nodos rama sucesivos hasta que x concuerde con una clave o se alcance una hoja. Es evidente que si una hoja etiquetada como K_i es alcanzada, esto significa que x es mayor que la clave K_i pero menor que la clave K_{i+1} (si se alcanza la hoja K_0, esto significa que x es menor que K_1. Si se alcanza la hoja K_n, significa que x es mayor que K_n). Por ejemplo, sean *AB, CF, EG, PP* las claves K_1 K_2, K_3 y K_4 en el árbol de búsqueda de la figura 6.15. Dado el objeto *BB*, los pasos de búsqueda de acuerdo con la figura 6.15 son:

1. Compare *BB* con K_3, la cual es *EG*.
2. Como *BB* es menor que *EG*, compare *BB* con K_1 la cual es *AB*.
3. Como *BB* es mayor que *AB*, compare *BB* con K_2, la cual es *CF*.
4. Como *BB* es menor que *CF*, se alcanza la hoja K_1.

Así, concluimos que el objeto *BB* es mayor que *AB* y menor que *CF*.

Un criterio obvio para medir la efectividad de un procedimiento de búsqueda es el número máximo de comparaciones que el procedimiento efectúa en el peor caso, esto es, la altura del árbol de búsqueda correspondiente. Por tanto, para un conjunto dado de n claves, un árbol de búsqueda cuya altura es $\lceil \lg (n+1) \rceil$ corresponderá a un procedimiento de búsqueda que posiblemente sea el mejor. Por otro lado, ya que siempre existe un árbol de

† En un árbol binario, el hijo izquierdo (derecho) de un nodo rama es la raíz del subárbol izquierdo (derecho) del nodo.

búsqueda binario de altura $\lceil \lg (n+1) \rceil$ para cualquier $n^{\dagger}$ el problema de diseñar un procedimiento de búsqueda que posiblemente sea el mejor de acuerdo con este criterio, no es un problema difícil.

En general, los resultados de nuestras búsquedas podrían ocurrir con diferentes frecuencias. Por ejemplo, en el problema de adivinar el número que un amigo nuestro ha seleccionado, si sabemos que es más probable que él seleccione un número menor que 20, que un número mayor o igual a 20, podríamos reconsiderar nuestra estrategia al hacer las preguntas. Para ser específicos, supongamos que de 1 000 búsquedas nos dan $u_1, u_2, \ldots, u_n$ como las frecuencias de aparición, los resultados de que un objeto dado sea igual a $K_1, K_2, \ldots, K_n$, y nos dan $w_0, w_1, w_2, \ldots, w_n$ como las frecuencias de aparición de los resultados de que un objeto dado sea menor que K_1, sea mayor que K_1 pero menor que K_2,. . ., sea mayor que K_n. Dado un árbol de búsqueda para las claves $K_1, K_2, \ldots, K_n$, el número total de comparaciones en 1 000 búsquedas será igual a

$$\sum_{j=1}^{n} u_j\,[l(K_j)+1] + \sum_{j=0}^{n} w_j\, l'(K_j) \tag{6.3}$$

donde $l(K_J)$ es la longitud del paseo del nodo rama etiquetado como K_j y $l'(K_j)$ es la longitud del paseo de la hoja etiquetada con K_j en el árbol de búsqueda. Observemos que $l(K_J) + 1$ es el número de comparaciones llevadas a cabo por el procedimiento de búsqueda si el objeto dado es igual a K_i, y $l'(K_i)$ es el número de comparaciones llevadas a cabo por el procedimiento de búsqueda si el objeto dado es mayor que K_j y menor que K_{j+1}.

El problema de construir un árbol de búsqueda para minimizar la cantidad en (6.3) para $u_1, u_2, \ldots, u_n, w_0, w_1, w_2, \ldots, w_n$ dados, fue analizado en Knuth [9]. Una variación del problema fue estudiada por Hu y Tucker [5] y Hu [4]. Entrar en los detalles de sus resultados estaría fuera del alcance de nuestro análisis. El lector interesado puede consultar estos artículos, así como el capítulo 6 de Knuth [8].

6.6 ÁRBOLES GENERADORES Y CONJUNTOS DE CORTE

Sea G un grafo conexo donde los vértices representan los edificios de una fábrica, y las aristas representan los túneles de conexión entre los edificios. Uno desearía determinar un subconjunto de túneles que debieran mantenerse abiertos todo el tiempo de manera que pudiéramos alcanzar un edificio desde otro a través de estos túneles. También desearíamos determinar los subconjuntos de túneles que al ser obstruidos, separarían a algunos de los edificios de los otros. En esta sección estudiamos los conceptos de *subconjuntos de aristas de conexión y subconjuntos de aristas de no conexión* en un grafo.

Un *árbol de un grafo* es un subgrafo del grafo que es un árbol. Un *árbol generador de un grafo conexo* es un subgrafo generador del grafo que es un árbol. Por ejemplo, la figura 6.16*b* muestra un árbol, y la figura 6.16*c* muestra un árbol generador del grafo de la figura 6.16*a*. Una *rama* de un árbol es una arista del grafo que es un árbol. Una *cuerda,* o un

† Véase el problema 6.15.

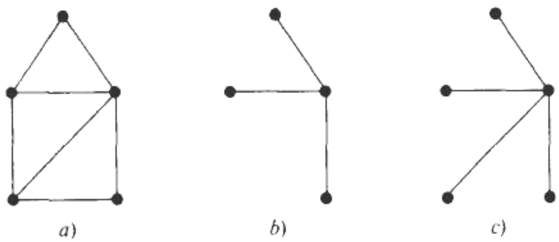


Figura 6.16

enlace de un árbol, es una arista del grafo que no está en el árbol. El conjunto de cuerdas de un árbol se conoce como el *complemento* del árbol.

Observamos que un *grafo conexo siempre contiene un árbol generador.* Supongamos que nos dan un grafo conexo. Si el grafo no contiene circuitos, entonces es un árbol. Si el grafo contiene uno o más circuitos, podemos eliminar una arista de uno de los circuitos y aún tener un subgrafo conexo. Dicha eliminación de aristas de los circuitos puede repetirse hasta que tengamos un árbol generador. De este argumento se obtiene de inmediato, que un árbol generador es un subgrafo conexo mínimo† de un grafo conexo en el sentido de que a partir de un subgrafo conexo el cual no es un árbol generador, una o más de sus aristas pueden eliminarse, de manera que el grafo resultante aún sea un subgrafo conexo, y, por otra parte, ninguna arista puede eliminarse de un árbol generador de manera que el subgrafo resultante aún sea un subgrafo conexo. En el ejemplo de los edificios conectados por túneles, si mantenemos abiertos los túneles correspondientes a las aristas en un árbol generador, nos aseguramos que podemos llegar desde un edificio a otro a través de estos túneles. Además, éste sería un conjunto mínimo de túneles que deben mantenerse abiertos.

Para un grafo conexo con e aristas y v vértices, existen v - 1 ramas en cualquier árbol generador. Entonces, en relación con cualquier árbol generador, existen e - v + 1 cuerdas.

Un *conjunto de corte* es un conjunto (mínimo) de aristas en un grafo tal que la eliminación del conjunto incrementará el número de componentes conexas en el subgrafo restante, en tanto que la eliminación de cualquier subconjunto propio de éste no lo haría. De esto se tiene que en un grafo conexo, la eliminación de un conjunto de corte divide al grafo en dos partes. Esto sugiere una forma alternativa de definir un conjunto de corte. Si los vértices de una componente conexa de un grafo se dividen en dos subconjuntos, de manera que cada dos vértices en cada subconjunto estén conectados por un paseo que sólo atraviesa vértices en tal subconjunto, entonces, el conjunto de aristas que une los vértices de los dos subconjuntos es un conjunto de corte. Como ejemplo, para el grafo de la figura 6.17*a*, el conjunto de aristas $\{e_1, e_5, e_6, e_7, e_4\}$ es un conjunto de corte, ya que su eliminación dejará un subgrafo no conexo como se muestra en la figura 6.176, en tanto que la eliminación de cualquiera de sus subconjuntos propios no lo hará. Además, éste es el conjunto de aristas que unen los vértices de los dos subconjuntos $\{v_1, v_5\}$ y $\{v_2, v_3, v_4\}$. La figura 6.17*a* se volvió a hacer en la figura 6.17*c* para enfatizar la división de vértices. En el ejemplo de los edificios conectados por túneles, si los túneles correspondientes a las aristas en un conjunto de corte

† Un *subgrafo conectado* de un grafo es un subgrafo generador que es conexo.

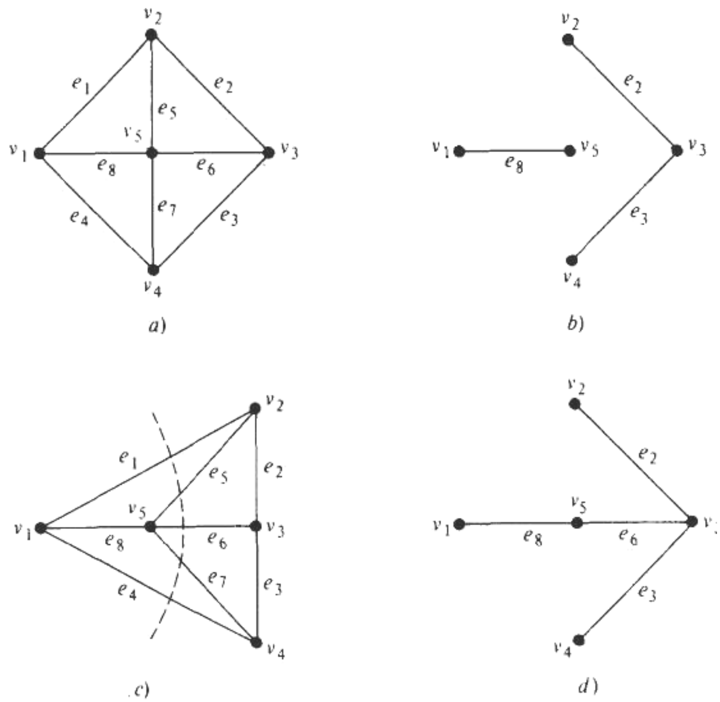


Figura 6.17

son bloqueados, entonces los edificios quedarán separados en dos conglomeraciones sin paso alguno desde un edificio en una conglomeración hacia un edificio en la otra.

Los conceptos de árboles generadores, circuitos y conjuntos de corte están muy relacionados. Debido a que un árbol generador contiene un paseo único entre dos vértices cualesquiera en el grafo, la adición de una cuerda al árbol generador da lugar a un subgrafo que contiene exactamente un circuito. Supongamos que se añade la cuerda $\{ v_l , v_2\}$ aun árbol generador. Como el árbol generador contiene un paseo entre v_l y v_2, este paseo, con la arista $\{ v_l, v_2\}$ forman un circuito en el grafo. Por otro lado, si la adición de la cuerda $\{v_1, v_2\}$ da origen a dos o más circuitos, deben existir dos o más paseos entre v_x y v_2 en el árbol generador, lo que obviamente es imposible. Para un árbol generador dado, se puede obtener un circuito único al agregar al árbol generador cada una de las cuerdas. El conjunto de $e - v + 1$ circuitos obtenidos de esta manera es llamado *el sistema fundamental de circuitos* relativo al árbol generador. Un circuito en el sistema fundamental es llamado un *circuito fundamental.* Puesto que un circuito fundamental contiene exactamente una cuerda del árbol generador, éste es conocido como el circuito fundamental correspondiente a la cuerda. Por ejemplo, para el grafo de la figura 6.17*a* y el árbol generador de la figura 6.17*d* el circuito fundamental correspondiente a la cuerda e_1, es el circuito $\{e_1, e_2, e_6, e_8\}$, y los otros circuitos fundamentales son $\{e_5, e_2, e_6\}$, $\{e_4, e_8, e_6, e_3\}$ y $\{e_7, e_6, e_3\}$.

Ya que la eliminación de cualquier rama de un árbol generador rompe el árbol generador en dos árboles (donde cualquiera de los dos o ambos pueden consistir en un vértice aislado), decimos que, correspondiente a una rama de un árbol generador, existe una división de los vértices del grafo en dos subconjuntos correspondientes a los vértices de los dos árboles.

Entonces, se tiene que para toda rama en un árbol generador, existe un correspondiente conjunto de corte. Por ejemplo, para el grafo de la figura 6.17*a*, la eliminación de la rama e_3 del árbol generador de la figura 6.17*d* divide a los vértices en dos subconjuntos $\{v_1, v_2, v_3, v_5\}$ y $\{v_4\}$. El conjunto de corte correspondiente es $\{e_3, e_7, e_4\}$. Para un árbol generador dado, el conjunto de los v - 1 conjuntos de corte correspondientes a las v - 1 ramas del árbol generador es llamado el *sistema fundamental de conjuntos de corte* relativo al árbol generador. Un conjunto de corte en el sistema fundamental de conjuntos de corte es llamado un *conjunto de corte fundamental.* Puesto que un conjunto de corte fundamental contiene exactamente una rama del árbol, éste es conocido como el conjunto de corte fundamental correspondiente a la rama. Para el grafo de la figura 6.17a y el árbol generador de la figura 6.17*d* los conjuntos de corte fundamentales son $\{e_1\ e_5, e_2\}$, $\{e_1, e_8, e_4\}$ $\{e_1, e_5, e_6, e_7, e_4\}$ y $\{e_4, e_7, e_3\}$.

Ahora presentamos algunas de las propiedades de circuitos y conjuntos de corte. A menos que se diga otra cosa, nuestro análisis se limitará a grafos conexos, ya que su extensión a grafos no conexos es directa.

Teorema 6.1

Un circuito y el complemento de cualquier árbol generador deben tener al menos una arista en común.

DEMOSTRACIÓN Si existe un circuito que no tenga una arista en común con el complemento de un árbol generador, el circuito está contenido en el árbol generador. Sin embargo, esto es imposible ya que un árbol no puede contener un circuito.

□

Teorema 6.2

Un conjunto de corte y cualquier árbol generador deben tener al menos una arista en común.

DEMOSTRACIÓN Si existe un conjunto de corte que no tenga una arista en común con un árbol generador, la eliminación del conjunto de corte dejará al árbol generador intacto. Sin embargo, esto significa que la eliminación del conjunto de corte no divide al grafo en dos componentes, lo cual contradice la definición de un conjunto de corte.

□

Teorema 6.3

Cada circuito tiene un número par de aristas en común con cada conjunto de corte.

DEMOSTRACIÓN Correspondiente a un conjunto de corte, existe una división de los vértices del grafo en dos subconjuntos, los cuales son los dos conjuntos de vértices de las dos componentes del grafo cuando las aristas del conjunto de corte son eliminadas. Por tanto, un paseo que conecta dos vértices en un subconjunto debe recorrer las aristas en el conjunto de corte un número par de veces, como se ilustra en la figura 6.18 (las aristas en el circuito son trazadas en líneas gruesas). Debido a que un circuito es un paseo desde algún vértice hasta él mismo, se cumple el teorema.

□

Subconjunto de vértices　　　Subconjunto de vértices

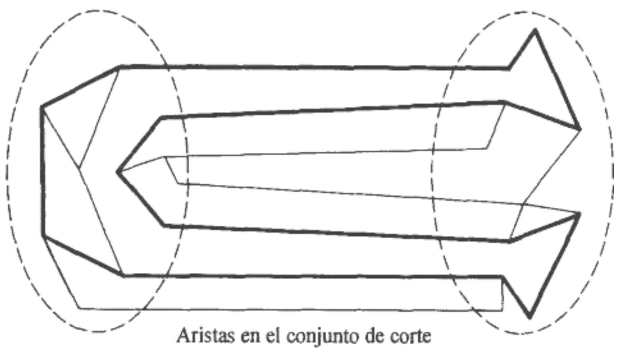


Figura 6.18

El siguiente resultado indica la gran relación entre el sistema fundamental de circuitos y el sistema fundamental de conjuntos de corte relativos a un árbol generador:

Teorema 6.4

Para un árbol generador dado, sea $D = \{e_1, e_2, e_3, \ldots, e_k\}$ un conjunto de corte fundamental, en el cual e_l es una rama y $e_2, e_3, \ldots, e_k$ son cuerdas del árbol generador, entonces e_x está contenido en los circuitos fundamentales correspondientes a e_x para $i = 2,3,\ldots,k$. Además, e_1 no está contenido en cualesquiera otros circuitos fundamentales.

DEMOSTRACIÓN Sea C el circuito fundamental correspondiente a la cuerda e_2. Observemos que e_2 está tanto en C como en D. Puesto que C y D tienen un número par de aristas en común y e_x es la única otra arista que posiblemente puede estar en C y en D† e_1 debe estar contenida en C. Un argumento similar puede aplicarse a los circuitos fundamentales correspondientes a las cuerdas $e_3, e_4, \ldots, e_k$. Por otro lado, si C' es el circuito fundamental correspondiente a cualquier cuerda que no está en D. C' no puede contener a e_1; porque en otro caso, C' y D tendrían a e_1 como la única arista en común.

❒

De modo similar tenemos:

Teorema 6.5

Para un árbol generador dado, sea $C = \{e_1\ e_2, e_3, \ldots, e_k\}$ un circuito fundamental en el cual e_1 es una cuerda y $e_2, e_3, \ldots, e_k$ son ramas del árbol generador, entonces e_1 está contenido en los conjuntos de corte fundamentales correspondientes a e_1 para $i = 2, 3, \ldots, k$. Además, e_1 no está contenido en cualesquiera otros conjuntos de corte fundamentales.

Dejamos la demostración del teorema 6.5 como un ejercicio (véase problema 6.21), ya que es prácticamente igual a la del teorema 6.4.

---

† Todas las aristas en C con excepción de e_2 son ramas, y todos las aristas en D con excepción de e, son cuerdas.

6.7 ÁRBOLES GENERADORES MÍNIMOS

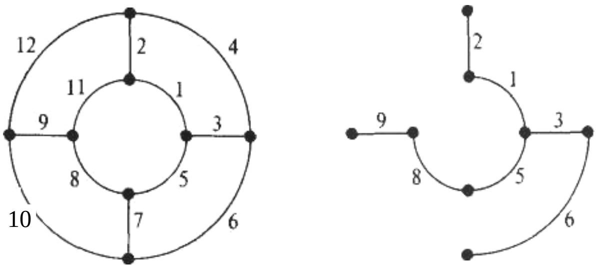
Para un grafo dado, uno podría querer determinar un árbol generador del grafo. Aquí consideramos el problema más general de determinar un árbol generador mínimo en un grafo pesado, donde a las aristas se les asignan números reales como sus pesos. El peso de un árbol generador se define como la suma de los pesos de las ramas del árbol. Un *árbol generador mínimo* es uno con peso mínimo. Una interpretación física de este problema es considerar los vértices de un grafo como ciudades, y los pesos de las aristas como los costos de construcción y mantenimiento de vías de comunicación entre las ciudades. Supongamos que queremos construir una red de comunicaciones que conecte a todas las ciudades a un costo mínimo. Entonces el problema es determinar un árbol generador mínimo. Presentamos dos algoritmos simples para realizar esto.

Nuestro primer procedimiento se basa en la observación de que entre todas las aristas en un circuito, la arista con el mayor peso no está en un árbol generador mínimo.† Sea *C* un circuito en un grafo pesado, y *e* la arista con el mayor peso en *C*. Supongamos que *e* es una rama de un árbol generador *T*. Sea *D* el conjunto de corte fundamental correspondiente a la rama *e*. Como el circuito *C* y el conjunto de corte *D* deben tener un número par de aristas en común, además de la arista *e* deberán existir al menos una o más aristas que estén tanto en C como en *D*. Sea *f* una de esas aristas. Observemos que *f* es una cuerda del árbol generador *T* debido a que *D* es un conjunto de corte fundamental. Agreguemos la arista *f* al árbol generador *T* y denotemos el subgrafo resultante como *U*. Es obvio que *U* es un subgrafo generador que contiene exactamente un circuito, el circuito fundamental correspondiente a *f*. De acuerdo con el teorema 6.4, e está contenido en el circuito fundamental correspondiente a *f* Si eliminamos *e* de *U*, obtenemos un árbol generador cuyo peso es menor que el de *T*.

Nuestra observación sugiere un procedimiento para determinar un árbol generador mínimo de un grafo pesado conexo. Construiremos un subgrafo del grafo pesado paso por paso, al ir examinando cada arista en orden creciente de los pesos. Se agregará una arista al subgrafo parcialmente construido si su inclusión no origina un circuito, y será descartada en caso contrario. La construcción termina cuando todas las aristas han sido examinadas. Es claro que nuestra construcción da origen a un subgrafo que no contiene un circuito. Observemos que el subgrafo también es conexo, ya que para una arista *{a, b}* en el grafo original, ya sea que la arista *{a, b}* esté incluida en el subgrafo o bien, existe un paseo entre *a* y *b* en el subgrafo. Así, el subgrafo que hemos construido es un árbol. Además, éste es un árbol generador debido a que el grafo original es conexo. Finalmente, el árbol generador es mínimo porque en el procedimiento de construcción una arista era excluida en favor de las aristas de pesos mayores sólo si se sabía que la arista excluida no podía estar en un árbol generador mínimo. En otras palabras, las v - 1 aristas en el subgrafo son efectivamente las v - 1 aristas con los pesos menores que pueden ser incluidas en un árbol generador mínimo.

---

† Para simplificar la presentación, suponemos que los pesos de las aristas son distintos. En el caso de que los pesos de las aristas no sean todos distintos, este resultado debe establecerse en una forma más general: de entre todas las aristas de un circuito, una arista con el mayor peso no está contenida en *algún* árbol generador mínimo.



a) *b)* **Figura 6.19**

Como ejemplo, un árbol generador mínimo para el grafo pesado de la figura 6.19*a* se muestra en la figura 6.19*b*. Observemos cómo las aristas con peso 4 y peso 7 son excluidas en la construcción paso por paso.

Un segundo procedimiento para construir un árbol generador mínimo se basa en la observación de que entre todas las aristas incidentes con un vértice, la arista con el menor peso debe estar en el árbol generador mínimo.† Sean v_1 un vértice y { v_1, v_2} la arista con el menor peso de entre todas los aristas incidentes con v_1. Sea *T* un árbol generador que no contiene la arista {$v_{1;}$ v_2}. Agreguemos la arista {$v_{];}$ v_2} a *T* y denotemos como *U* al subgrafo resultante. Señalemos que *U* contiene exactamente un circuito, el circuito fundamental correspondiente a la cuerda { v_1, v_2}. Este circuito está formado por la arista *{v_h v_2} y* el paseo de v_1 hasta $\underline{v}_2$ en *T*. Sea $(v_1, v_{i_1}, v_{i_2}, \ldots, v_{i_k}, v_2)$ la sucesión de vértices de tal paseo. Observamos que al eliminar $\{v_1, v_{i_1}\}$ de *U*, obtenemos un árbol generador cuyo peso es menor que el de *T*.

Sean *G = (V, E)* un grafo, y v_1 y v_2 dos vértices de *V*. Introducimos la noción de obtener un grafo *G'* a partir de *G* mediante la *fusión* de los vértices v_1 y v_2. Intuitivamente, *G'* es un grafo obtenido a partir de *G* al combinar los vértices v_1 y v_2 en un "supervértice", y al retener todas las aristas en *G*. Por ejemplo, la figura 6.20¿ muestra un grafo obtenido a partir del grafo de la figura 6.20*a* mediante la fusión de los vértices v_1 y v_2. Debido a que *G y G'* son, en general, multigrafos, identificaremos las aristas de *E* mediante nombres de aristas, como e_1 y e_2, en lugar de los vértices con los cuales ellos son incidentes. Sea *G'* = (*V*, *E*') tal que *V´* contiene todos los vértices en *V*, excepto que los vértices v_1 y v_2 son eliminados y un nuevo vértice *v** es introducido, y *E'* contiene todas las aristas de *E*, excepto que si una arista fuera incidente con v_1 o v_2 en *G*, éste es incidente con v* en *G'*‡ Tenemos la siguiente observación: sea *e* una arista con el menor peso que no es un lazo en *G*. Sea *G'* el grafo obtenido a partir de *G* mediante la fusión de los vértices v_1 y v_2 y con los cuales la arista *e* es incidente en *G*, y *T'* un árbol generador mínimo de *G'*. Denotemos como *T* a un subgrafo

† De nuevo, suponemos que los pesos de las aristas son distintos. En el caso de que los pesos de las aristas no sean todos distintos, este resultado debe establecerse en una forma más general: de entre todas las aristas incidentes con un vértice, una arista con el menor peso está contenida en *algún* árbol generador mínimo.

‡ La arista entre v_1 y v_2 se convierte en un lazo en *v** (con el objetivo de construir un árbol generador mínimo, podríamos optar por eliminar a todos los lazos en *G'*. Sin embargo, no hacemos esto porque no queremos desviarnos de la definición convencional de fusión de dos vértices en un grafo).

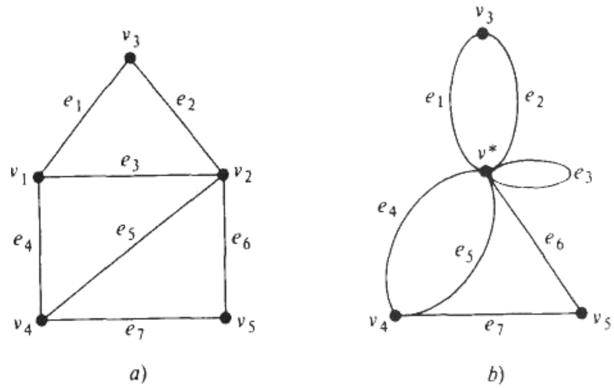


Figura 6.20

de *G* que consiste en todas las aristas de *T'* y la arista *e*. Queremos mostrar que *T* es un árbol generador mínimo de *G*. Primero, señalemos que *T* efectivamente es un árbol generador de *G*. En segundo lugar tenemos

$$W(T) = W(T') + w\{e)$$

donde *w(e)* denota el peso de la arista *e*, y *W(T) y W(T')* denotan la suma de los pesos de las aristas de *T y T'*, respectivamente. Por último, si *T* no fuera un árbol generador mínimo de *G,* existe un árbol generador mínimo de *G,* $\hat{T},$ el cual contiene la arista *e* de manera que

$$W(\hat{T},)<W(T)$$

Sea $\hat{T}'$ el árbol obtenido de $\hat{T}$ mediante la fusión de los vértices v_1 y v_2 y la eliminación de la arista *e*. Es claro que, *T'* es un árbol generador de *G'*. Puesto que

$$W(\hat{T}) = W(\hat{T}') + w(e)$$

obtenemos

$$W(\hat{T}') < W(T')$$

lo cual contradice la suposición de que *T´* es un árbol generador mínimo del grafo *G'*.

Nuestras observaciones sugieren otro procedimiento para determinar un árbol generador mínimo en un grafo pesado conexo. Sea $\{ v_1, v_2 \}$ la arista de menor peso en un grafo *G*. Como $\{ v_1, v_2 \}$ debe incluirse en un árbol generador mínimo de *G*, podemos fusionar los dos vértices v_1 y v_2 para obtener el grafo *G*', y entonces determinar un árbol generador mínimo de *G'*. El paso puede repetirse hasta que terminemos con un grafo que tenga un solo vértice.

Como ejemplo, para el grafo pesado de la figura 6.19a, se muestran varios pasos de la construcción en las figuras 6.21*a* a 6.21*b*. Observemos que las aristas con los pesos 4 y 7 son excluidas en la construcción paso por paso.

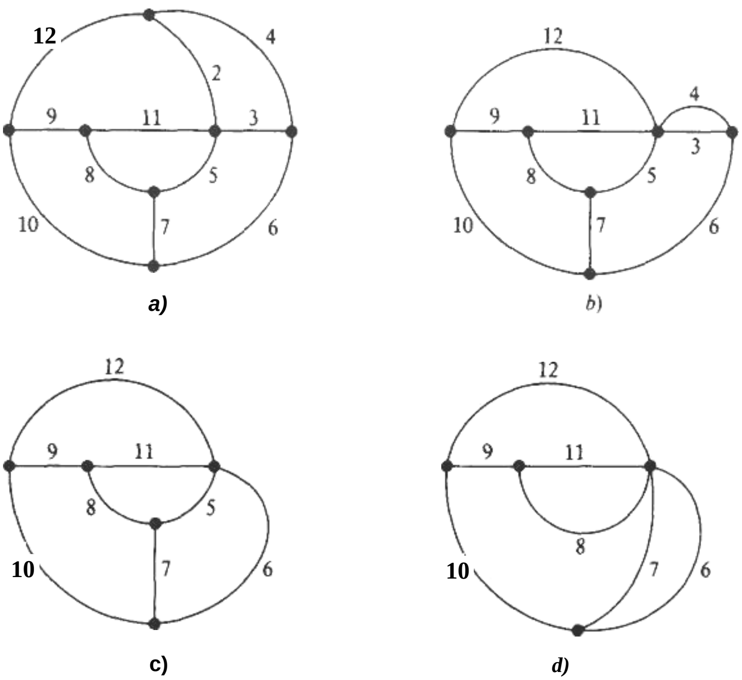


Figura 6.21

*6.8 REDES DE TRANSPORTE

Diremos que un grafo dirigido pesado es una *red de transporte* si se satisfacen las siguientes condiciones:

1. El grafo es conexo y no tiene lazos.
2. Existe uno y sólo un vértice en el grafo que no tiene una arista que llegue a él.
3. Existe uno y sólo un vértice en el grafo que no tiene una arista que salga de él.
4. El peso de cada arista es un número real no negativo.

En una red de transporte, el vértice que no tiene aristas que llegan a él, es llamado la *fuente* y es denotado por *a;* el vértice que no tiene aristas que salgan de él es llamado el *sumidero* y se denota por *z*. El peso de una arista se llama la *capacidad* de la arista. La capacidad de la arista *(i, j)* se denota por *w(i,j)*.

Una red de transporte representa un modelo general para el transporte de material desde el origen de los suministros hasta el destino a través de rutas de envío, donde existen límites superiores sobre la cantidad de material que puede ser enviado a través de las rutas. La figura 6.22*a* muestra un ejemplo de una red de transporte.

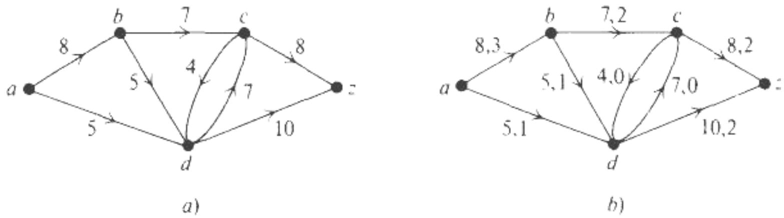


Figura 6.22

Un *flujo* en una red de transporte, ϕ, es una asignación de números no negativos $\phi(i, j)$ a cada arista (i, j), de manera que se satisfacen las siguientes condiciones:

1. $\phi(i, j) \leqq w(i, j)$ para cada arista (i, j).
2. $\sum_{\text{toda } i} \phi(i, j) = \sum_{\text{toda } k} \phi(j, k)$ para cada vérticey excepto la fuente a y el sumidero z .[†]

En términos de transporte de material, $\phi(i, j)$ es la cantidad de material a ser enviado a través de la ruta (*i, j*). La condición 1 significa que la cantidad de material a ser enviada a través de una ruta, no debe exceder la capacidad de la ruta. La condición 2 significa que, excepto en la fuente y el sumidero, la cantidad de material que fluye hacia un vértice debe ser igual a la cantidad de material que fluye hacia afuera del vértice. Por ejemplo, la figura 6.22¿ muestra un flujo en la red de transporte de la figura 6.22*a*. El primer número asociado con una arista es la capacidad de la arista, y el segundo número asociado con una arista es el flujo de la arista. La cantidad $\sum_{\text{toda } i} \phi(a, i)$ se dice que es el *valor del flujo* ϕ y se denota por ϕ_v. Intuitivamente, es claro que

$$\phi_v = \sum_{\text{toda } i} \phi(a, i) = \sum_{\text{toda } k} \phi(k, z)$$

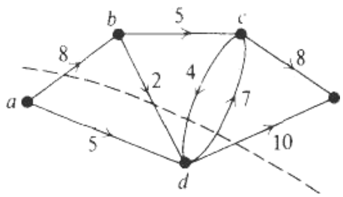
es decir, el flujo total que sale de la fuente es igual al flujo total que entra en el sumidero. Este resultado se demuestra rigurosamente en la demostración del teorema 6.6. Para un flujo dado, diremos que una arista (i, j) está *saturada* si $\phi(i, j) = w(i, j)$ y se dice que *no está saturada* si $\phi(i, j) < w(i, j)$. Un *flujo máximo* en una red de transporte es el fluio que alcanza el valor más grande posible. Es concebible que pueda existir más de un flujo máximo en una red de transporte. En otras palabras, puede haber varios flujos diferentes, los cuales todos alcanzan el valor más grande posible.

Con frecuencia queremos determinar la mayor cantidad de material que puede ser enviado desde la fuente hasta el sumidero de una red de transporte dada. Por ello, es deseable tener un algoritmo para construir un flujo en la red que alcance el mayor valor posible. Para realizar esto presentaremos un algoritmo, pero primero introduciremos algunos conceptos y resultados útiles.

Un *corte* en una red de transporte es un conjunto de corte del grafo no dirigido, obtenido de la red de transporte al ignorar la dirección de las aristas, que separa la fuente del sumidero.

---

† Definimos $\phi(x, y)$ como cero si no existe una arista desde x a y.



(P, P) **Figura 6.23**

La notación $(P, \overline{P})$ se usa para denotar un corte que divide los vértices en dos subconjuntos *P* y *P*, donde el subconjunto *P* contiene a la fuente y el subconjunto *P* contiene al sumidero. La *capacidad de un corte*, denotada por $w(P, \overline{P})$, se define como la suma de las capacidades de aquellas aristas incidentes de los vértices de *P* hasta los vértices de *P;* esto es,

$$w(P, \overline{P}) = \sum_{i \in P, j \in \overline{P}} w(i, j)$$

Por ejemplo, la línea punteada en la figura 6.23 identifica un corte que separa el subconjunto de vértices $P = \{a, d\}$ del subconjunto de vértices $\overline{P} = \{b, c, z)$. La capacidad de este corte es igual a 8+ 7+ 10 = 25.

El siguiente resultado da una cota superior para los valores de flujos en una red de transporte.

Teorema 6.6 El valor de un flujo cualquiera en una red de transporte dada es menor o igual que la capacidad de cualquier corte en la red.

DEMOSTRACIÓN Sea ϕ un flujo y $(P, \overline{P})$ un corte en una red de transporte. Para la fuente *a*,

$$\sum_{\text{toda } i} \phi(a, i) - \sum_{\text{toda } j} \phi(j, a) = \sum_{\text{toda } i} \phi(a, i) = \phi_v \tag{6.3}$$

puesto que $\phi(j, a) = 0$ para todo j. Para un vértice p diferente de a en P,

$$\sum_{\text{toda } i} \phi(p, i) - \sum_{\text{toda } j} \phi(j, p) = 0 \tag{6.4}$$

Combinando (6.3) y (6.4), tenemos

$$\begin{aligned}
\phi_v &= \sum_{p \in P} \left[\sum_{\text{toda } i} \phi(p, i) - \sum_{\text{toda } j} \phi(j, p) \right] \\
&= \sum_{p \in P;\ \text{toda } i} \phi(p, i) - \sum_{p \in P;\ \text{toda } i} \phi(j, p) \\
&= \sum_{p \in P;\ i \in P} \phi(p, i) + \sum_{p \in P;\ j \in \overline{P}} \phi(p, i) \\
&\quad - \left[\sum_{p \in P;\ j \in P} \phi(j, p) + \sum_{p \in P;\ j \in \overline{P}} \phi(j, p) \right]
\end{aligned} \tag{6.5}$$

Observemos que

$$\sum_{p \in P;\, i \in P} \phi(p, i) = \sum_{p \in P;\, j \in P} \phi(j, p)$$

debido a que ambas sumas corren sobre todos los vértices de P. Así, (6.5) queda como

$$\phi = \sum_{p \in P;\, i \in \bar{P}} \phi(p, i) - \sum_{p \in P;\, j \in \bar{P}} \phi(j, p) \tag{6.6}$$

$$\phi_v \leqq \sum_{p \in P;\, i \in \bar{P}} \phi(p, i) \leqq \sum_{p \in P;\, i \in \bar{P}} w(p, i) = w(P, \bar{P})$$

❒

Pero, ya que $\sum_{p \in P;\, j \in \bar{P}} \phi(j, p)$ siempre es una cantidad no negativa, tenemos

La ecuación (6.6) es un resultado útil, el cual puede establecerse como: para cualquier corte $(P, \bar{P})$, el valor de un flujo en una red de transporte es igual a la suma de los flujos de las aristas de los vértices de P hasta los vértices de $\bar{P}$ menos la suma de los flujos en las aristas de los vértices de $\bar{P}$ hasta los vértices de P.

Ahora estamos listos para presentar un algoritmo para construir un flujo máximo en una red de transporte. En vista del resultado de que el valor de cualquier flujo en una red de transporte es menor o igual a la capacidad de cualquier corte, siempre que podamos construir un flujo ϕ, el valor del cual es igual a la capacidad de un corte $(P, \bar{P})$, podemos estar seguros de que ϕ es un flujo máximo, porque si hubiera un flujo mayor, su valor excedería la capacidad del corte $(P, \bar{P})$. Entonces, siempre podremos construir un flujo cuyo valor sea igual a la capacidad de un corte con el siguiente procedimiento, conocido como el *procedimiento de etiquetado.* Para iniciar este procedimiento, debemos construir un flujo inicial ϕ en la red. No obstante, tal construcción no representa un problema puesto que siempre podemos iniciar, trivialmente, con un flujo de cero en cada arista.

Al inicio, la fuente a es estiquetada como $(-, \infty)$ (el significado de esta etiqueta se aclarará más adelante). A continuación, todos los vértices que son adyacentes desde a son barridos. Un vértice b que es adyacente desde a es etiquetado como $(a^+, \Delta b)$, donde Δb es igual a $w(a, b) - \phi(a, b)$, si $w(a, b) > \phi(a, b)$; no es etiquetado si $w(a, b) = \phi(a, b)$. Después de que todos los vértices adyacentes desde a han sido barridos y etiquetados (si es posible), aquellos vértices que son adyacentes hacia o desde los vértices etiquetados son barridos. Sea b un vértice etiquetado y sea q un vértice adyacente *desde b.* El vértice q es etiquetado como $(b^+, \Delta q)$, donde Δq es igual a la menor de las dos cantidades Δb y $[w(b, q) - \phi(b, q)]$ si $w(b, q) > \phi(b, q)$. El vértice q no es etiquetado si $w(b, q) = \phi(b, q)$. Sea b un vértice etiquetado, y sea q un vértice adyacente hacia b. El vértice q es etiquetado como $(b^-, \Delta q)$, donde Δq es igual a la menor de las dos cantidades Δb y $\phi(q, b)$ si $\phi(q, b) > 0$. El vértice q no es etiquetado si $\varphi(q, b) = 0$. Este procedimiento de etiquetado no necesariamente es único. El vértice q podría ser adyacente hacia o desde más de un vértice etiquetado. Además, podría existir una arista incidente desde b hacia q también como una arista incidente desde q hacia b. En cualquier caso, cuando un vértice puede etiquetarse en más de una manera, se realiza una selección arbitraria de alguna de estas maneras.

Examinemos los significados de estas etiquetas antes de proceder con la presentación de los pasos restantes del procedimiento. Para un vértice que es adyacente desde la fuente (como el vértice *b),* la etiqueta $(a^+, \Delta b)$ significa que el flujo hacia *b* puede incrementarse por una cantidad igual a Δb. Además, tal incremento puede obtenerse desde la fuente *a.* De modo similar, para un vértice *q* que es adyacente desde un vértice etiquetado *b,* la etiqueta $(b^+, \Delta q)$ significa que, al obtener el incremento desde el vértice *b,* el flujo total que entra a *q* desde los vértices etiquetados puede incrementarse una cantidad Δq. Para un vértice *q* que es adyacente a un vértice etiquetado *b,* la etiqueta $(b^-, \Delta q)$ significa que, al disminuir el flujo
desde *q* hacia *b,* el flujo total que sale desde *q* hacia los vértices etiquetados puede disminuirse en una cantidad Δq. En cualquiera de estos casos, se asegura un incremento en el flujo igual a Δq desde el vértice *q* hacia los vértices no etiquetados. El significado de la etiqueta de la fuente, $(-, \infty)$ deberá aclararse en este momento. Esto significa que (partiendo de aquí hacia
afuera) la fuente puede suministrar una cantidad infinita de material hacia los otros vértices.

Si repetimos el procedimiento de etiquetar los vértices que son adyacentes hacia o desde los vértices etiquetados, surgirá uno de los dos siguientes casos:

Caso 1 El sumidero z está etiquetado, digamos, con una etiqueta $(y^+, \Delta z)$ [desde luego, z nunca tendrá una etiqueta como $(y^-, \Delta z)$]. Podemos aumentar el flujo de la arista (y, z) desde $\phi(y, z)$ hasta $\phi(y, z) + \Delta z$, ya que el incremento está garantizado por el vértice *y*. Señalemos que el vértice *y* debe ser etiquetado como $(q^+, \Delta y)$ o como $(q^-, \Delta y)$, con $\Delta y \gtreqqless \Delta z$, para algún vértice *q*. Si *y* es etiquetado con $(q^+, \Delta y)$, obtendremos, a su vez, el incremento desde el vértice *q* al incrementar el flujo en la arista (q, y) desde $\phi(q, y)$ hasta $\phi(q, y) + \Delta z$. Por otro lado, si *y* es etiquetado con $(q^-, \Delta y)$, disminuiremos el flujo en la arista *(y, q)* desde $\phi(y, q)$ hasta $\phi(y, q) - \Delta z$, de manera que el incremento Δz de y a z se compensa. El proceso se lleva hasta la fuente *a, y* el valor del flujo en la red de transporte se incrementa una cantidad Δz. El procedimiento de etiquetado podrá iniciarse de nuevo para incrementar aún más el valor del flujo en la red.

Caso 2 El sumidero z no está etiquetado. Denotemos todos los vértices etiquetados por *P* y todos los vértices no etiquetados por $\bar{P}$. El hecho de que el sumidero z no esté etiquetado significa que el flujo en cada una de las aristas incidentes desde los vértices en *P* hasta los vértices en $\bar{P}$, es igual a la capacidad de tal arista, y que el flujo en cada una de las aristas incidentes desde los vértices de $\bar{P}$ hasta los vértices de *P* es igual a cero. Así, hemos obtenido un flujo, cuyo valor es igual a la capacidad del corte $(P, \bar{P})$. El flujo, luego entonces, es un flujo máximo.

Consideremos el siguiente ejemplo. Para la red de transporte de la figura 6.24*a*, iniciamos con un flujo de cero en cada arista (el primer número asociado con una arista es su capacidad, y el segundo número es el flujo de la arista). La figura 6.24*b* muestra el primer paso del procedimiento de etiquetado.Observemos que el sumidero z puede ser etiquetado ya sea con $(d^+, 3)$, o con $(b^+, 2)$. Escogemos arbitrariamente la etiqueta $\{d^+, 3)$. La figura 6.24*c* muestra el segundo paso, y la figura 6.24*d* el tercer paso. En la figura *6.24d* el vértice *b* es etiquetado con $(c^+, 6)$ y el vértice *d* es etiquetado con $(c^+, 4)$; esto es, el vértice *c* tiene garantizado un flujo total de 10 hacia los vértices *b* y *d,* aun cuando Δc sólo es igual a 9. No

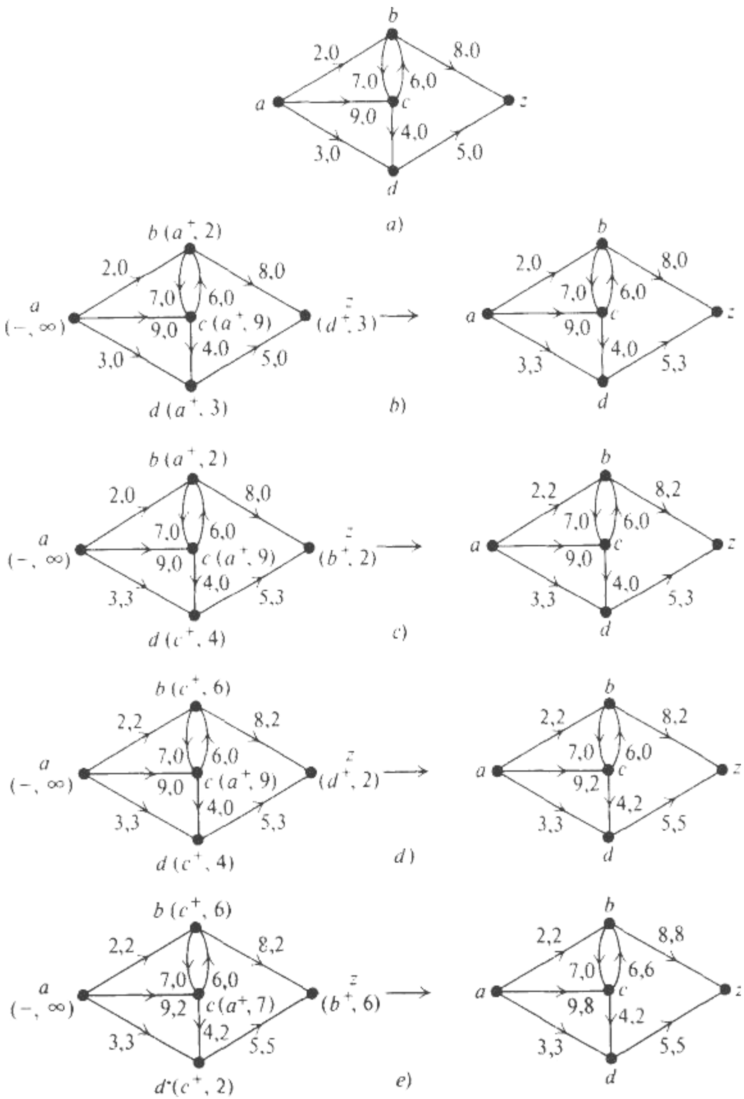


Figura 6.24

obstante, ya que el vértice *c* podría proporcionar un suministro hacia el incremento del flujo en *b*, o hacia el incremento del flujo en *d*, pero no en ambos, en el paso de aumento no se tiene ninguna dificultad. La figura 6.24*e* muestra el último paso del procedimiento de etiquetado, el cual origina un flujo máximo de 13.

Como otro ejemplo, consideremos la red de transporte de la figura 6.25*a*, donde se halla un flujo inicial. La figura 6.25*b* muestra cómo obtenemos un flujo máximo mediante el procedimiento de etiquetado.

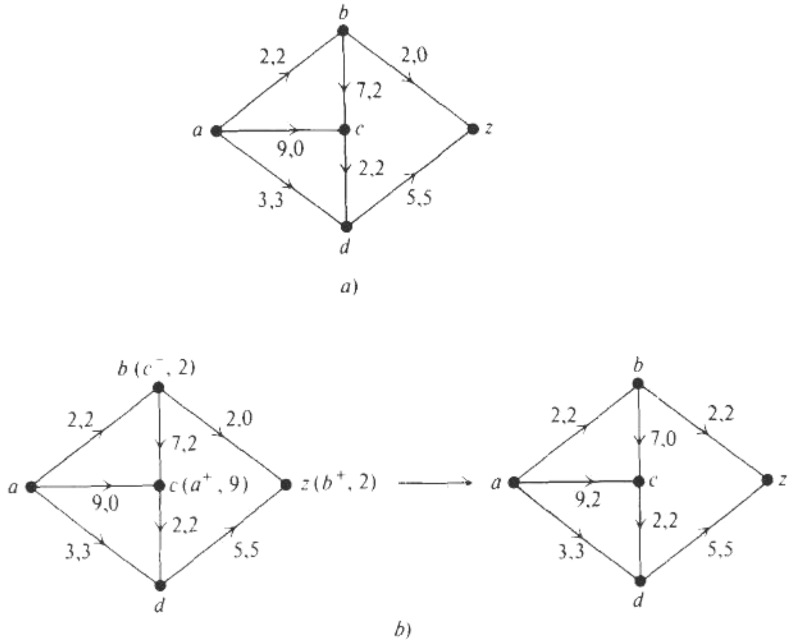


Figura 6.25

6.9 NOTAS Y REFERENCIAS

Véase en el capítulo 2 de Knuth [7] y el capítulo 6 de Knuth [8] un tratamiento completo sobre árboles y árboles de búsqueda. Véase Ford y Fulkerson [1], Hu [2], Hu [3], Lawler [10], Papadimitriou y Steiglitz [12], y Syslo, Deo y Kowalik [13] para más detalles sobre problemas de flujo en redes. De hecho; existen procedimientos "más eficientes" para encontrar un flujo máximo en una red de transporte. Escogimos el presentado en la sección 6.8 no sólo por ser un resultado clásico, sino también porque ilustra con claridad cómo podemos resolver un problema de optimización discreta en la modalidad de paso por paso; esto es, siempre mejoraremos la solución encontrada en cada paso y, más aún, existe un criterio que indica que quizá hemos alcanzado una mejor solución y que podemos detener nuestro procedimiento de mejorar paso a paso. Véase en el capítulo 7 de Liu [11] un análisis más extenso sobre circuitos y conjuntos de corte.

1. Ford, L. R., Jr., y D. R. Fulkerson: *Flows in Networks,* Princeton University Press, Princeton, N.J., 1962.
2. Hu, T. C: *Combinatorial Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1982.
3. Hu, T. C: *Integer Programming and Network Flows,* Addison-Wesley Publishing Company, Reading, Mass., 1969.
4. Hu, T. C: "A New Proof of the T-C Algorithm", *SIAM Journal on Applied Mathematics,* **25:** 83-94(1971).
5. Hu, T. C, y A. C. Tucker: "Optimum Computer Search Trees", *SIAM Journal on Applied Mathematics,* 21: 514-532 (1971).

6. Huffman, D. A.: "A Method for the Construction of Minimum Redundancy Codes", *Proc. IRÉ,* 40:1098-1101(1952).
7. Knuth, D. E.: *The Art of Computer Programming, Vol. 1, Fundamental Algorithms,* 2d ed., Addison-Wesley Publishing Company, Reading, Mass., 1973.
8. Knuth, D. E.: *The Art ofComputer Programming, Vol. 3, Sorting andSearching,* Addison-Wesley Publishing Company, Reading, Mass., 1973.
9. Knuth, D. E.:" Optimum Binary Search Trees", *Acta Informática,* 1: 14-25 (1971).
10. Lawler, E.: *Combinatorial Optimization,* Holt, Rinehart and Winston, Nueva York, 1976.
11. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hül Book Company, Nueva York, 1968.
12. Papadimitriou, C. H., y K. Steiglitz: *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall, Englewood Cliffs, N.J.,1982.
13. Syslo, M. M., N. Deo, y J. S. Kowalik: *Discrete Optimization Algorithms,* Prentice-Hall, Englewood Cliffs, N.J., 1983.
14. Zipf, G. K.: *Human Behavior and the Principie of Least Ejfort, an Introduction to Human Ecology,* Addison-Wesley Publishing Company, Reading, Mass., 1949.

PROBLEMAS

6.1 Determine todos los árboles que tienen exactamente dos hojas.

6.2 Un árbol tiene $2n$ vértices de grado 1, $3n$ vértices de grado 2, y n vértices de grado 3. Determine el número de vértices y aristas en el árbol.

6.3 *a)* Un árbol tiene dos vértices de grado 2, un vértice de grado 3 y tres vértices de grado 4. ¿Cuántos vértices de grado 1 tendrá el árbol?

b) Un árbol tiene n_2 vértices de grado 2, n_3 vértices de grado 3, . . . , y n_k vértices de grado k. ¿Cuántos vértices de grado 1 tendrá el árbol?

6.4 Sea T un árbol con 50 aristas. La eliminación de una cierta arista de T origina dos árboles disjuntos T_1 y T_2- Puesto que el número de aristas en T_1 es igual al número de aristas en T_2, determine el número de vértices y el número de aristas en T_1 y en T_2.

6.5 *a)* Demuestre que la suma de los grados de los vértices de un árbol con n vértices es $2n - 2$.

b) Para $n \geqq 2$, sean $d_1, d_2, \ldots, d_n$ enteros positivos tales que $\sum_{i=1}^{n} d_i = 2n - 2$. Muestre que existe un árbol cuyos vértices tienen grados $d_1, d_2, \ldots, d_n$.

6.6 El centro de un grafo (conexo) se define como el vértice v con la propiedad de que la distancia máxima entre v y cualquier otro vértice es tan pequeña como sea posible (véase el problema 5.16 para la definición de la distancia entre dos vértices de un grafo).

a) Muestre un ejemplo de un grafo que tenga un centro.

b) Muestre un ejemplo de un grafo que tenga dos o más centros.

c) Muestre que un árbol tiene bien uno o dos centros. Más aún, si existen dos centros éstos deben ser adyacentes. Muestre un ejemplo de un árbol que tenga dos centros.

6.7 Sea v un vértice en un árbol T. Se puede asignar dirección a las aristas de T para obtener un árbol enraizado, siendo v la raíz. Denotemos por $v_1, v_2, \ldots, v_k$ a los hijos de v, y denotemos con $s_1, s_2, \ldots, s_k$ al número de vértices en los subárboles con $v_1, v_2, \ldots, v_k$ como raíces. El peso de v se define como el máximo de $s_1\ s_2, \ldots, s_k$.

a) Para el árbol de la figura 6P.1, calcule los pesos de todos los vértices.

b) Diremos que un vértice que tiene un peso mínimo es un *centroide* del árbol. Dé un ejemplo de un árbol que tenga un centroide, y un ejemplo de un árbol que tenga dos centroides.

c) Demuestre que un árbol tiene uno o dos centroides; si existen dos centroides, éstos deben ser adyacentes.

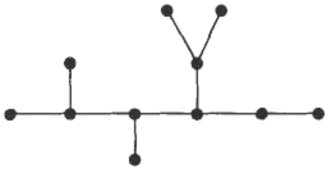


Figura 6P.1

d) Demuestre que v es el único centroide de un árbol si

$$s_j \leqq s_1 + s_2 + \cdots + s_k - s_j \qquad \text{para toda } j,\ 1 \leqq j \leqq k$$

e) Demuestre que un árbol con dos centroides tiene un número par de vértices. Además, el peso de cada centroide es igual a la mitad del número de vértices en el árbol.

6.8 Demuestre que un árbol binario regular tiene un número impar de vértices.

6.9 Demuestre el resultado de la ecuación (6.2) mediante inducción sobre *i*.

6.10 Por recorrer un árbol, entendemos visitar cada uno de los vértices del árbol exactamente una vez en algún orden secuencial. Describimos aquí tres de las principales maneras de recorrer un árbol binario.

1. *Recorrido preordenado:* visitamos la raíz, recorremos el subárbol izquierdo, y entonces recorremos el subárbol derecho.
2. *Recorrido en orden (orden simétrico):* recorremos el subárbol izquierdo, visitamos la raíz, y entonces recorremos el subárbol derecho.
3. *Recorrido postordenado:* recorremos el subárbol izquierdo, recorremos el subárbol derecho, y entonces visitamos la raíz.

Muestre los órdenes secuenciales en los cuales se visitan los vértices del árbol de la figura 6P.2 en un recorrido preordenado, en un recorrido en orden, y en un recorrido postordenado.

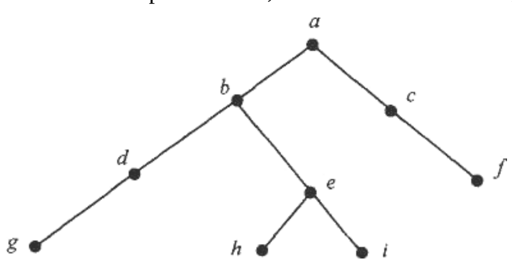


Figura 6P.2

6.11 Sea A un conjunto de sucesiones binarias. Particione A en dos subconjuntos A_0 y A_1 donde A_0 es el conjunto de sucesiones en A cuyo primer dígito es un 0 y A_1 es el conj unto de sucesiones en A cuyo primer dígito es un 1. Luego particione A_0 en dos subconjuntos de acuerdo con el segundo dígito de las sucesiones, y también A_1 de la misma manera. Utilice esta idea de particionar repetidamente un conjunto de sucesiones en subconjuntos para mostrar que si A es un código de prefijos, entonces existe un árbol binario, con las dos aristas incidentes desde cada uno de los nodos rama etiquetados con 0 y 1, de manera que las sucesiones de números 0 y 1 asignadas a las hojas son las sucesiones en A.

6.12 Por una *lista clasificada* de números entendemos una lista de números arreglados en orden ascendente. Por la *mezcla* de dos listas clasificadas entendemos el combinarlas en una sola lista clasificada. Describimos una manera para mezclar dos listas clasificadas: ya que el más pequeño de los números menores de las dos listas debe ser el menor de todos los números, podemos quitar este número de la lista en que está, y colocarlo en algún otro lugar como el primer número de la lista mezclada. Ahora podemos comparar los números menores de las dos listas de números restantes, y colocar al más pequeño de los dos como el segundo número de la lista mezclada. Este paso puede repetirse hasta que la lista mezclada esté completamente construida. Es obvio que esto tomará $n_1 + n_2 - 1$ comparaciones para mezclar dos listas clasificadas con n_1 y n_2 números, respectivamente. Dadas m listas clasificadas, podemos seleccionar dos de ellas y mezclar estas dos listas en una. Entonces seleccionamos dos listas de las m - 1 listas clasificadas y las mezclamos en una. Mediante la repetición de este paso, al final terminaremos con una lista mezclada. Sean A_1, A_2, A_3 y A_4 cuatro listas con 73, 44, 100 y 55 números, respectivamente.

a) Determine el número total de comparaciones que tomará mezclar las cuatro listas mezclando A_1 y A_2, mezclando la lista resultante con A_3, y luego mezclando la lista resultante con A_4.

b) Determine el número total de comparaciones que tomará mezclar las cuatro listas mezclando A_1 y A_2, mezclando A_3 y A_4, y luego mezclando las dos listas resultantes.

c) Determine un orden para mezclar las cuatro listas de manera que el número total de comparaciones sea un mínimo.

d) Describa un procedimiento general para determinar el orden en el cual m listas clasificadas $A_1, A_2, \ldots, A_m$ se mezclen de manera que el número total de comparaciones sea mínimo (*Sugerencia:* ¿cuál es la manera conveniente de describir cierto orden en el cual se mezclen las listas?).

6.13 Para cada uno de los siguientes conjuntos de pesos, construya un código de prefijos binarios óptimo. Para cada peso en el conjunto, dé la palabra código correspondiente.

a) 8,9,12,14,16,19.

b) 1,2,4,5,6,9,10,12.

c) 5,7,8,15,35,40.

6.14 *a)* ¿Cómo puede extenderse el procedimiento para la construcción de un árbol binario óptimo de la sección 6.4, de manera que se construya un árbol m-ario óptimo?

b) Construya un árbol ternario óptimo para los pesos 1, 2, 3, 4, 5, 6, 7, 8, 9.

c) Construya un árbol ternario óptimo para los pesos 1, 2, 3, 4, 5, 6, 7, 8. En general, cuando t - 1, donde t es el número de pesos, no es un múltiplo de m - 1 [véase ecuación (6.2)], ¿cómo deberemos proceder para construir un árbol m-ario óptimo?

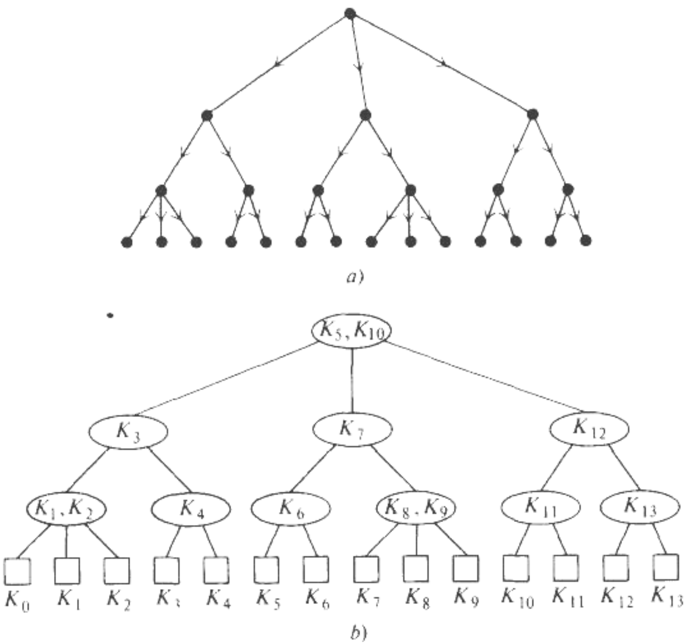


Figura 6P.3

6.15 Muestre cómo puede construirse un árbol de búsqueda binario de altura $\lceil \lg(n+1) \rceil$ para *n* claves, $K_1, K_2, \ldots, K_n$.

6.16 Definimos un *árbol 3-2* como un árbol enraizado en el cual el grado de salida de un nodo rama es 3 o 2. Además, las longitudes del paseo de todas las hojas deben ser las mismas. Por ejemplo, la figura 6P.3*a* muestra un árbol 3 - 2. Un árbol 3 - 2 puede usarse para describir una clase de procedimientos de búsqueda en los cuales comparamos un objeto con una o dos claves en cada paso de la búsqueda. Para el árbol 3 - 2 de la figura 6P.3*b* como un ejemplo, explique en detalle cómo trabaja este procedimiento de búsqueda. ¿Cómo podemos estar seguros de que el número de hojas en un árbol 3 - 2 siempre es igual al número de claves más uno?

6.17 Demuestre que el complemento de un árbol generador no contiene un conjunto de corte y que el complemento de un conjunto de corte no contiene un árbol generador.

6.18 Sea *L* un circuito en un grafo *G*. Sean *a* y *b* dos aristas cualesquiera en *L*. Demuestre que existe un conjunto de corte *C* tal que $L \cap C = \{a, b\}$.

6.19 Sean T_1 y T_2 dos árboles generadores de un grafo conexo *G*. Sea *a* una arista que está en T_1 pero no en T_2. Demuestre que existe una arista *b* en T_2 pero no en 7", tal que tanto $(T_1 - \{a\}) \cup \{b\}$ como $(T_2 - \{b\}) \cup \{a\}$ son árboles generadores de *G*.

6.20 *a)* Sean L_1 y L_2 dos circuitos en un grafo *G*. Sea *a* una arista que está tanto en L_1 como en L_2, y sea *b* una arista que está en L_1 pero no en L_2. Demuestre que existe un circuito L_3 tal que $L_3 \subseteq (L_1 \cup L_2) - \{a\}$ y $b \in L_3$.

b) Repita el inciso *a)* remplazando el término *circuito* por el término *conjunto de corte*.

6.21 Demuestre el teorema 6.5.

6.22 En este problema mostramos que existen n^{n-2} árboles generadores en un grafo completo con *n* vértices diferentemente etiquetados. Sin pérdida de generalidad, consideremos que las etiquetas de los vértices son $1, 2, \ldots, n$. Demostraremos una correspondencia uno a uno entre los árboles generadores y las $n"^{-2}$ $n - 2$ sucesiones de dígitos sobre el alfabeto $\{1, 2, \ldots, n\}$.Sea *T* un árbol generador cuyos vértices están etiquetados como $1, 2, \ldots, n$. Construimos una sucesión $a_1 a_2 \ldots a_{n-2}$ de la siguiente manera:

1. Sea $i = 1$. Sea *T* el árbol actualmente bajo examinación.
2. De entre todos los vértices de grado 1 en el árbol que se examina, seleccionamos uno con la etiqueta más pequeña. Eliminamos la arista que es incidente con este vértice y hacemos a_i igual a la etiqueta del *otro* vértice con el cual esta arista es incidente.
3. El árbol resultante en (2) es ahora el árbol que se examina. Incrementamos *i* por 1, y repetimos (2) hasta que se haya formado una sucesión de $n - 2$ dígitos.

a) Determine las sucesiones de 4 - dígitos correspondientes al árbol de la figura 6P.4.

b) Demuestre que el grado de un vértice con la etiqueta *i* en *T* es igual al número de veces que la etiqueta *i* aparece en $a_1 a_2 \ldots a_{n-2}$ más uno.

c) Determine un algoritmo para reconstruir un árbol a partir de una sucesión $a_1 a_2 \ldots a_{n-2}$.

d) Demuestre que el árbol reconstruido mediante el algoritmo del inciso c) es el único árbol que generará la sucesión $a_1 a_2 \ldots a_{n-2}$ de acuerdo con el procedimiento anteriormente descrito.

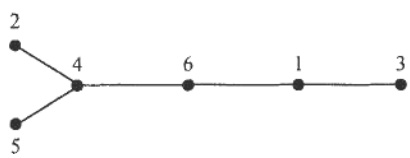


Figura 6P.4

6.23 *a)* Demuestre que cualquier arista de un grafo conexo *G* es una rama del algún árbol generador de *G*.

b) ¿Es verdad que cualquier arista de un grafo conexo *G* es una cuerda de algún árbol generador de *G*?

6.24 Sean G_1 y G_2 dos árboles. Sean v_1 y v_2 dos vértices distintos en G_1 y v_3 y v_4 dos vértices distintos en G_2. Sea G un grafo obtenido a partir de G_1 y G_2 al conectar v, con v_1 y v_2 con v_4, como se muestra en la figura 6P.5.

a) ¿Es G un árbol? Demuestre la respuesta.

b) ¿Es G un grafo conexo? Demuestre la respuesta.

c) ¿Qué puede decir acerca de G_1 y G_2 si resulta que G tiene un circuito euleriano?

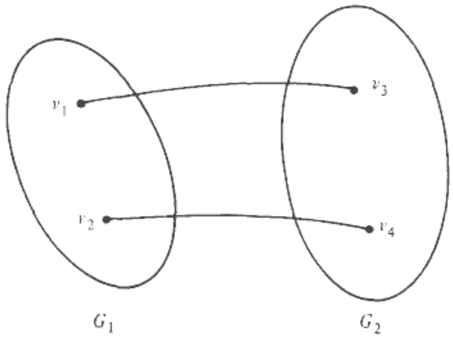


Figura 6P.5

6.25 ¿Cuántos árboles generadores diferentes hay en cada uno de los grafos de la figura 6P. 6? (Observe que los vértices son distintos).

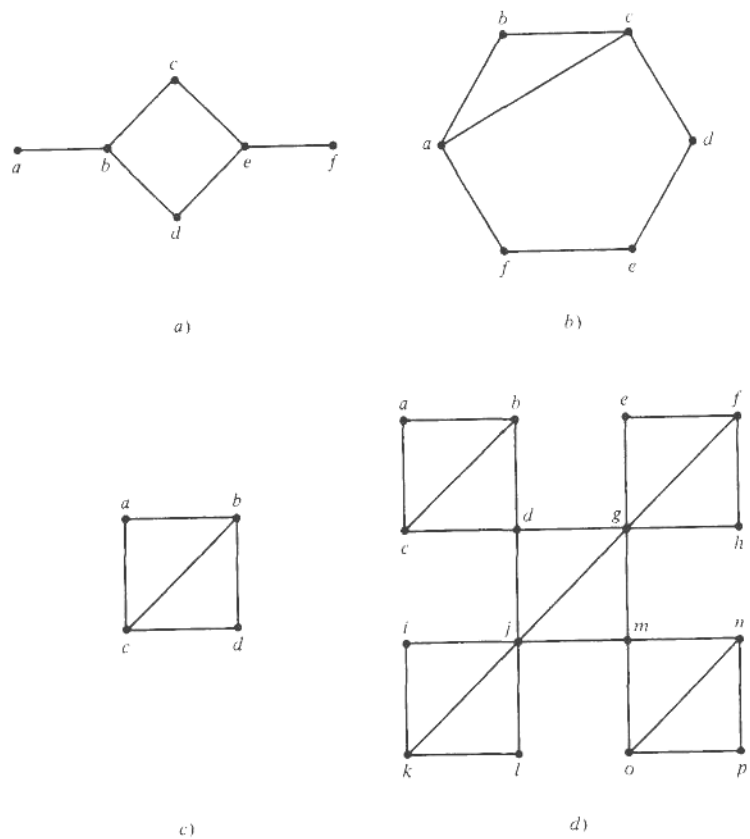


Figura 6P.6

6.26 Determine un árbol generador mínimo para el grafo mostrado en la figura 6P.7.

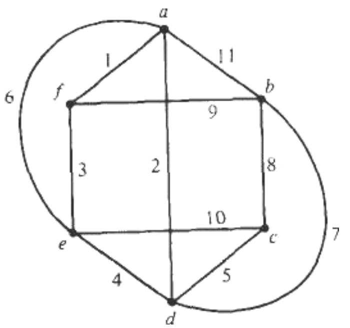


Figura 6P.7

6.27 Determine un árbol generador mínimo para el grafo mostrado en la figura 6P.8.

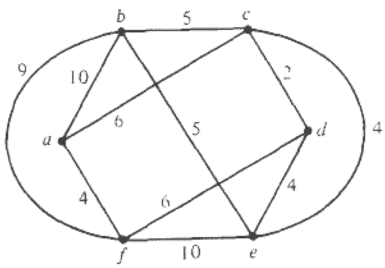


Figura 6P.8

6.28 Determine un árbol generador mínimo para el grafo mostrado en la figura 6P.9.

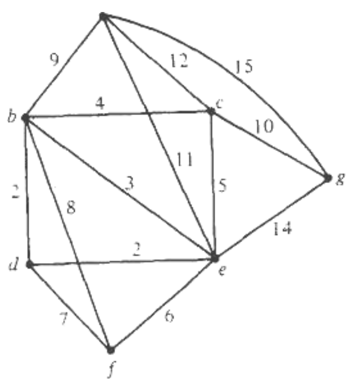


Figura 6P.9

6.29 Determine un árbol generador mínimo para el grafo mostrado en la figura 6P. 10.

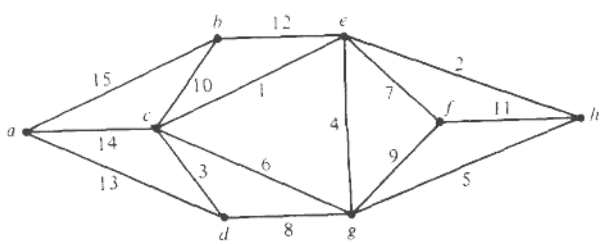


Figura 6P.10

6.30 Determine un árbol generador mínimo para el grafo mostrado en la figura 6P.11

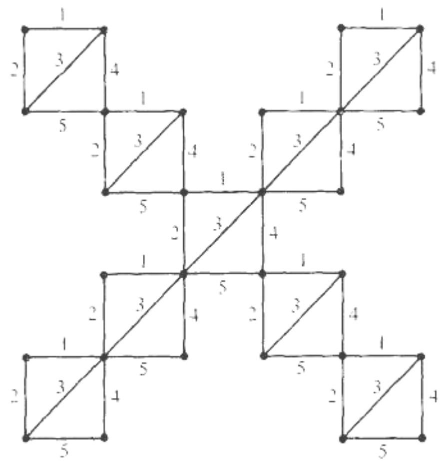


Figura 6P.11

6.31 La compañía de teléfonos cometió un error y tendió demasiadas líneas telefónicas entre un grupo de casas. En el grafo mostrado en la figura 6P.12, los vértices son las casas y las aristas son las líneas telefónicas. Las longitudes de las aristas son las longitudes de la líneas. Para subsanar el problema, la compañía de teléfonos quiere eliminar las líneas telefónicas extra, de manera que la suma de las longitudes de las líneas restantes sea la más pequeña posible, sujetos a la condición de que cada casa está conectada a cualquier otra casa mediante un paseo de líneas telefónicas. Córtense las líneas que deben ser eliminadas y determine la longitud total de las líneas restantes.

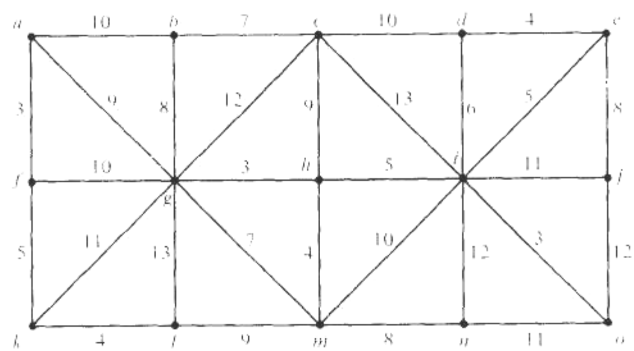


Figura 6P.12

6.32 Utilice el procedimiento de etiquetado para encontrar un flujo máximo en la red de transporte de la figura 6P.13. Determine el correspondiente corte mínimo.

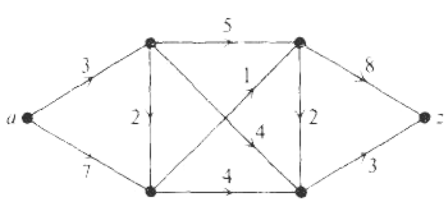


Figura 6P.13

6.33 Utilice el procedimiento de etiquetado para encontrar un flujo máximo en la red de transporte de la figura 6P.14. Determine el correspondiente corte mínimo.

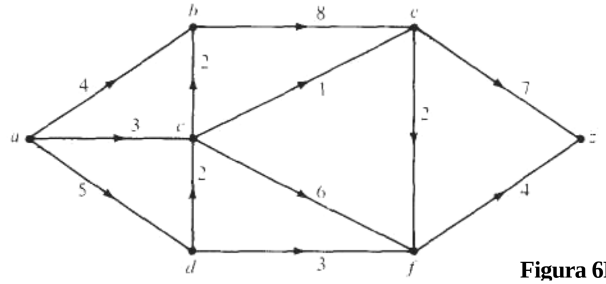


Figura 6P.14

6.34 Utilice el procedimiento de etiquetado para encontrar un flujo máximo en la red de transporte de la figura 6P. 15. Determine el correspondiente corte mínimo.

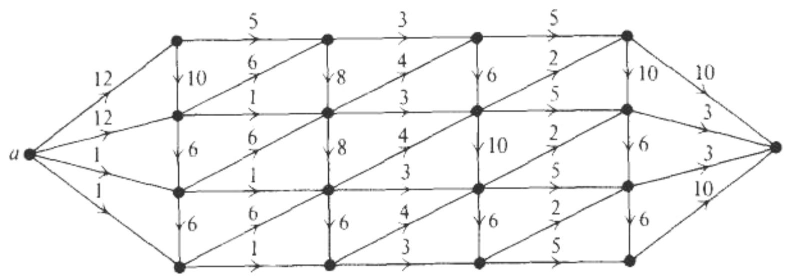


Figura 6P.15

6.35 Cierto equipo es manufacturado en tres fábricas x_1, x_2 y x_2, y es enviado a tres depósitos _y_1 y_2 y y_3 a través de la red de transporte mostrada en la figura 6P. 16.

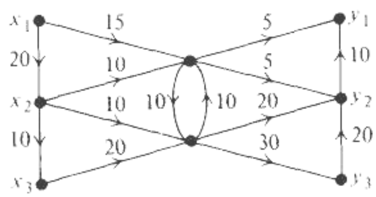


Figura 6P.16

La fábrica x_1 puede producir 40 unidades, la fábrica x_2 puede producir 20 unidades, y la fábrica x_3 puede producir 10 unidades. El depósito y_1 necesita 15 unidades, el depósito y_2 necesita 25 unidades, y el depósito y_3 necesita 10 unidades. ¿Cuántas unidades deberá producir cada fábrica de manera que éstas puedan ser transportadas a los depósitos?

6.36 Encuentre un flujo máximo en la red de transporte de la figura 6P.17 en la cual el flujo de las aristas no orientadas puede realizarse en cualquier dirección.

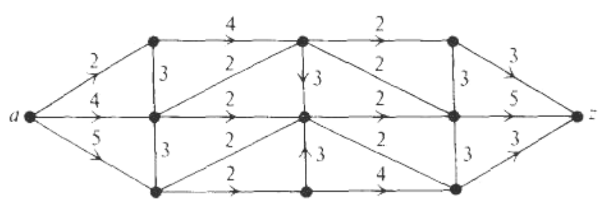


Figura 6P.17

6.37 *a)* Siete tipos de equipo militar serán enviados a un determinado destino mediante cinco aeroplanos de carga. Hay cuatro unidades de cada tipo, y los cinco aeroplanos pueden cargar ocho, ocho, cinco, cuatro y cuatro unidades, respectivamente. ¿Puede cargarse el equipo de manera que no haya dos unidades del mismo tipo en un aeroplano?

b) Dé una solución al inciso *a)* cuando las capacidades de los aeroplanos son siete, siete, seis, cuatro y cuatro unidades, respectivamente.

6.38 Construya un grafo dirigido con cuatro vértices y con no más de dos aristas desde un vértice a otro de manera que los grados de salida y de entrada de los vértices sean (5, 4), (3, 3), (1, 2) y (2, 2), respectivamente, mediante la correspondiente solución al problema de flujo en la red.

6.39 Dada una red de transporte, deseamos encontrar un flujo tal que el flujo en cada arista sea mayor o igual que la capacidad de la arista y el valor del flujo es mínimo.

á) Defina la "capacidad de un corte", y establezca la condición de flujo-mínimo corte-máximo, la cual es análoga a la condición de flujo-máximo corte-mínimo presentada en la sección 6.8.

b) Demuestre la condición de flujo-mínimo corte-máximo mediante el diseño de un algoritmo para encontrar un flujo mínimo.

6.40 Ingenieros y técnicos serán contratados por una compañía para participar en tres proyectos. Los requerimientos para el personal de estos tres proyectos se listan en la siguiente tabla:

	Número mínimo de personas necesarias en cada proyecto	Número mínimo en cada categoría			
		Ingenieros mecánicos	Técnicos mecánicos	Ingenieros eléctricos	Técnicos eléctricos
Proyecto I	40	5	10	10	5
Proyecto II	40	10	5	15	5
Proyecto III	40	5	0	10	5

Además, preparándose para una futura expansión, la compañía quiere contratar al menos 30 ingenieros mecánicos, 20 técnicos mecánicos, 20 ingenieros eléctricos y 20 técnicos electricistas. ¿Cuál es el número mínimo de personas de cada categoría que deberá contratar la compañía? y ¿cómo deberá distribuirlas en los tres proyectos?

6.41 En el grafo de la figura 6P.18, un conjunto mínimo de aristas será seleccionado de manera que cada vértice sea incidente con al menos una de las aristas en el conjunto. Solucione este problema como un problema de flujo-mínimo asociado con una red de transporte.

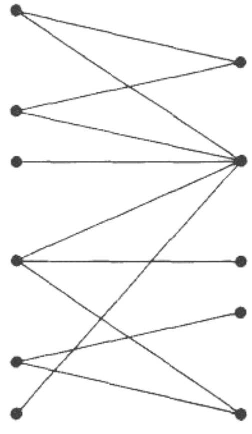


Figura 6P.18

6.42 A los tres candidatos x_1 x_2 y x_3 les han sido prometidas para su campaña, cantidades mínimas de dinero de \$40 000, \$23 000 y \$50 000, respectivamente. Cada candidato, a su vez, ha prometido tres áreas de campaña con al menos las cantidades de dinero mostradas en la tabla siguiente. Además, cada candidato necesitará al menos \$5 000 para sus propios gastos.

	Área de campaña		
Candidato	C_1	C_2	C_3
x_1	\$20 000	\$10 000	\$10 000
x_2	10 000	5 000	2 000
x_3	5 000	10 000	20 000

Si las tres áreas de campaña C_1 C_2 y C_3 requieren un mínimo de \$30 000, \$25 000 y \$50 000, respectivamente, para llevar a cabo la campaña completa, ¿cuál es una cantidad mínima de dinero para campaña que debe obtener cada candidato? y, ¿cómo deberá ser distribuida?

CAPÍTULO

SIETE

Máquinas de estado finito

7.1 INTRODUCCIÓN

Una *máquina procesadora de información* es un dispositivo que recibe un conjunto de señales de entrada, y produce en correspondencia un conjunto de señales de salida, como se ilustra en la figura 7.1. Por tanto, podemos considerar que una lámpara de mesa es una máquina procesadora de información: la señal de entrada es la posición *ENCENDIDO* o la *posición APAGADO* del interruptor, y la señal de salida es la *ILUMINACIÓN* o la *OSCURIDAD*. Otro ejemplo de máquina procesadora de información es un sumador, cuyas señales de entrada son dos números decimales y la señal de salida su suma. En un automóvil, que es una máquina procesadora de información, las señales de entrada son la presión sobre el acelerador y la posición angular del volante, y las señales de salida corresponden a la velocidad y dirección del vehículo. En una máquina vendedora, también una máquina procesadora de información las señales de entrada son las monedas depositadas y la selección de la mercancía, las señales de salida son la mercancía y, posiblemente, el cambio. Por último, una computadora digital es una máquina procesadora de información; el programa del usuario y los datos son las señales de entrada, y los resultados impresos del cálculo son las señales de salida.

En general, las señales de entrada para una máquina procesadora de información cambian con el tiempo. En ese caso, las señales de salida también variarán con el tiempo, 'en la forma correspondiente. Así, una máquina procesadora de información recibe una serie (temporal) de señales de entrada y produce una correspondiente serie (temporal) de señales de salida. Consideremos el ejemplo de una lámpara de mesa: la señal de entrada es una de las dos posibles posiciones del interruptor, *ENCENDIDO* o *APAGADO,* y la señal de salida es

Figura 7.1

una de las dos posibles condiciones, *ILUMINACIÓN* u *OSCURIDAD*. Por tanto, en correspondencia con la serie de señales de entrada

ENCENDIDO APAGADO APAGADO ENCENDIDO
APAGADO ENCENDIDO ENCENDIDO . . .

existe la serie de señales de salida

ILUMINACIÓN OSCURIDAD OSCURIDAD ILUMINACIÓN
OSCURIDAD ILUMINACIÓN ILUMINACIÓN . . .

En el ejemplo del sumador, donde las señales de entrada son dos números de un dígito y la señal de salida es un número de dos dígitos, a las series de señales de entrada son

$$\begin{array}{cccccccc} 3 & 5 & 0 & 3 & 3 & 9 & 2 & \ldots \\ 4 & 4 & 6 & 1 & 4 & 5 & 5 & \ldots \end{array}$$

y la correspondiente serie de señales de salida es

$$\begin{array}{cccccccc} 7 & 9 & 6 & 4 & 7 & 14 & 7 & \ldots \end{array}$$

Consideremos el ejemplo de la máquina vendedora, donde las señales de entrada son monedas de cinco, diez o veinticinco centavos, y la señal de salida es un paquete de goma de mascar o nada. Suponga que un paquete de goma de mascar cuesta 30 centavos. Entonces, como corresponde a la serie de señales de entrada

DIEZ DIEZ DIEZ VEINTICINCO VEINTICINCO
CINCO VEINTICINCO CINCO . . .

existe la serie de señales de salida†

NADA NADA GOMA DE MASCAR NADA GOMA DE MASCAR
NADA GOMA DE MASCAR NADA . . .

† Suponemos que la máquina vendedora no regresa cambio.

Advirtamos que existe una diferencia significativa entre las máquinas de estos ejemplos. En el caso de la lámpara de mesa, siempre que la señal de entrada es *ENCENDIDO,* la señal de salida es *ILUMINACIÓN,* y siempre que la señal de entrada es *APAGADO,* la señal de salida es *OSCURIDAD.* Es decir, la señal de salida en depende en cualquier momento sólo de la señal de entrada activada en tal instante, y no de las señales de entrada anteriores a dicho instante. En el caso del sumador sucede lo mismo, la señal de salida en cualquier momento siempre es la suma de los dos números dados como entrada en tal instante, y es independiente por completo de los números que fueron sumados con anterioridad. Por otro lado, en el caso de la máquina vendedora, la señal de salida obtenida en cualquier instante depende no sólo de la señal de entrada dada en tal instante, sino, además, de las señales de entrada precedentes. Así, para las tres señales de entrada sucesivas

DIEZ DIEZ DIEZ

las correspondientes señales de salida son

NADA NADA GOMA DE MASCAR

En forma más específica, en el primer instante la entrada es *DIEZ* y la correspondiente salida es *NADA;* en el segundo instante la entrada es *DIEZ* y la salida correspondiente es *NADA,* pero en el tercer instante, la entrada es *DIEZ* y la salida correspondiente es *GOMA DE MASCAR.* Es obvio que esta observación no sorprende a nadie, porque todos estamos enterados de que una máquina vendedora es capaz de recordar la cantidad total que ha sido depositada, mientras que una lámpara de mesa no es capaz (o no necesita ser capaz) de recordar las señales de entrada previas.

Es por esto que dividimos a las máquinas en dos clases -unas con memoria y otras sin memoria. Para una máquina sin memoria su salida en cualquier instante, sólo depende de la entrada en tal instante. Tanto la lámpara de mesa como el sumador analizados, son ejemplos de máquinas que no tienen memoria. Para una máquina con memoria, su salida en cualquier instante depende de la entrada en dicho instante como de las entradas en instantes previos debido a que la máquina puede recordar "qué ha sucedido en el pasado". Es claro que una máquina vendedora puede recordar qué ha sucedido en el pasado. No obstante, ésta no recuerda (o no puede recordar) *todo* lo que ha sucedido en el pasado. En un instante, la máquina recuerda la cantidad total que ha sido depositada hasta entonces. Es decir, mientras la cantidad total-sea, digamos, 25 centavos, la máquina no hace diferencia alguna acerca de si se depositaron cinco monedas de cinco centavos, dos monedas de diez centavos y una de cinco centavos, una moneda de veinticinco centavos, u otras combinaciones posibles. Para describir los pasados eventos, introducimos el concepto de *estado.* Un estado representa un resumen de la historia pasada de una máquina. En el ejemplo de la máquina vendedora existen siete estados diferentes, que corresponden al depósito total acumulado -a saber, 0, 5, 10, 15, 20, 25 y 30 o más centavos. En consecuencia, el estado de la máquina junto con las señales de entrada en un instante particular, determinarán las señales de salida correspondientes a dicho instante. Siguiendo con el ejemplo de esta máquina, en un instante cualquiera, el estado en que se encuentra la máquina más el nuevo depósito, permitirán a la máquina determinar si deberá tener *NADA* o *GOMA DE MASCAR* como salida. Además, conforme llega otra señal de entrada, la máquina pasará desde un estado a otro, puesto que necesita actualizar el resumen de su historia. Es decir, debe actualizar la cantidad total que ha sido depositada. Así, por ejemplo, cuando la máquina se encuentra en el estado de 15 cen-

Depósito total	Depósito nuevo		
	5¢	10¢	25¢
0¢	5¢	10¢	25
5¢	10¢	15¢	30¢ o más
10¢	15¢	20¢	30¢ o más
15¢	20¢	25¢	30¢ o más
20¢	25¢	30¢ o más	30¢ o más
25¢	30¢ o más	30¢ o más	30¢ o más
30¢ o más	5¢	10¢	25¢

Depósito total nuevo

a)

Depósito total	Mercancía entregada
0¢	Nada
5¢	Nada
10¢	Nada
15¢	Nada
20¢	Nada
25¢	Nada
30¢ o más	Goma de mascar

b)

Figura 7.2

tavos, una entrada de 10 centavos llevaría a la máquina al estado de 25 centavos o una entrada de 25 centavos la llevaría al estado de 30 centavos o más, etcétera. El comportamiento de la máquina vendedora puede resumirse como se muestra en la figura 7.2 (las notaciones de la figura 7.2, aunque son obvias, serán introducidas formalmente en la sección 7.2).

Como otro ejemplo, consideremos una máquina que acepta una sucesión de enteros positivos entre 1 y 100, y produce como salida en cualquier instante el entero más grande que ha recibido hasta ese momento, como se ilustra en la figura 7.3. Observemos que en tanto la máquina pueda recordar el mayor entero que ha recibido, cuando llega una nueva entrada, la máquina puede comparar el mayor entero recibido hasta el momento con la nueva entrada, y determinar la salida correspondiente, la cual es el "nuevo" entero más grande recibido hasta el momento. Así, para esta máquina, un resumen de la historia pasada puede ser representado mediante un entero igual al mayor entero que haya recibido. Por tanto, la máquina deberá tener 101 estados correspondientes a los enteros 0, 1, . . . , 100, que representan el mayor entero que la máquina ha recibido (es claro que el estado 0 significa

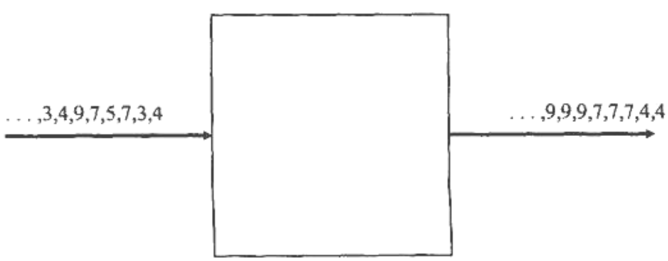


Figura 7.3

que ningún entero ha sido recibido). Desde luego, uno podría diseñar una máquina que recuerde el mayor y el segundo más grande enteros recibidos. El lector podría estar de acuerdo con que a pesar de que la máquina se comportaría correctamente, ésta mantendría alguna información redundante (el segundo entero más grande recibido), lo cual no contribuye a la correcta operación de la máquina.

Una máquina puede tener un cierto número de estados correspondientes a un cierto número de clases distintas de historias pasadas. Una máquina con un número finito de estados se llama *máquina de estado finito.* Por otro lado, una máquina con un número infinito de estados se llama *máquina de estado infinito.* En este capítulo restringiremos nuestro estudio a máquinas de estado finito.

7.2 MÁQUINAS DE ESTADO FINITO

Ahora introducimos un modelo abstracto de una máquina de estado finito, la cual está especificada por:

1. Un conjunto finito de estados $S = \{s_0, s_1, s_2, \ldots\}$.
2. Un elemento especial del conjunto S, s_0, es conocido como *estado inicial.*
3. Un conjunto finito de caracteres de entrada $I = \{i_1, i_2, \ldots\}$.
4. Un conjunto finito de caracteres de salida $O = \{o_1, o_2, \ldots\}$.
5. Una función *f* de *S x I* a *S,* conocida como *función de transición.*
6. Una función *g* de *S* a *O*, conocida como *Junción de salida.*

En un instante cualquiera, una máquina de estado finito se encuentra en uno de sus estados. Al llegar un carácter de entrada, la máquina pasará a otro estado de acuerdo con la función de transición. Además, en cada estado la máquina produce un carácter de salida de acuerdo con la función de salida. Al principio, la máquina se encuentra en su estado inicial. Una forma conveniente para describir una máquina de estado finito es la forma tabular usada en la figura 7.4. Para la máquina de estado finito de la figura 7.4, el conjunto de estados es $\{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$, el conjunto de caracteres de entrada es $\{a, b, c\}$, y el conjunto de caracteres de salida es $\{0, 1\}$. La doble flecha que apunta hacia s_0 indica que s_0 es el estado inicial. La función de transición/está especificada en la figura 7.4a, donde el estado en la intersección de una fila (correspondiente al estado s_p) y una columna (correspondiente al

	Estado	Entrada		
		a	b	c
⇒	s_0	s_1	s_2	s_5
	s_1	s_2	s_3	s_6
	s_2	s_3	s_4	s_6
	s_3	s_4	s_5	s_6
	s_4	s_5	s_6	s_6
	s_5	s_6	s_6	s_6
	s_6	s_1	s_2	s_5

a)

Estado	Salida
s_0	0
s_1	0
s_2	0
s_3	0
s_4	0
s_5	0
s_6	I

b)

Figura 7.4

	Estado	Entrada			Salida
		a	b	c	
$\Rightarrow$	s_0	s_1	s_2	s_5	0
	s_1	s_2	s_3	s_6	0
	s_2	s_3	s_4	s_6	0
	s_3	s_4	s_5	s_6	0
	s_4	s_5	s_6	s_6	0
	s_5	s_6	s_6	s_6	0
	s_6	s_1	s_2	s_5	1

Figura 7.5

carácter de entrada i_q), es el valor $f(s_p, i_q)$† La función de salida se especifica en la figura *lAb*. Por lo regular, las figuras 7.4a y 7.46 pueden combinarse en una sola tabla, como en la figura 7.5. De hecho, la máquina de estado finito de la figura 7.4 es precisamente la máquina vendedora mostrada de la figura 7.2. En específico, los estados $s_0, s_1, s_2, s_3, s_4, s_5$ y s_6, corresponden a los estados 0,5,10,15,20,25 y 30 o más centavos. Los caracteres de

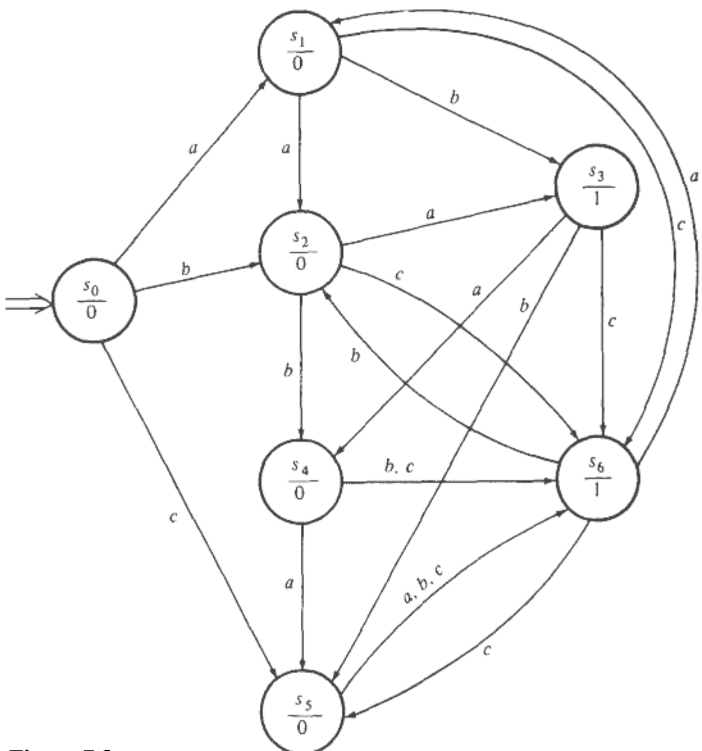


Figura 7.6

†Para ser exactos, la notación debe ser, $f((s_p, i_q))$. No obstante, seguimos la práctica común de omitir un par de paréntesis dado que no existe confusión como ya señalamos en el capítulo 4.

entrada *a*, *b* y c corresponden a las monedas de *CINCO, DIEZ* y *VEINTICINCO* centavos. Los caracteres de salida 0 y 1 corresponden al momento en que la máquina entregará -*NADA* o *GOMA DE MASCAR.*

Podemos describir una máquina de estado finito en forma gráfica, como se muestra en la figura 7.6. En el grafo dirigido de la figura 7.6, cada vértice corresponde a un estado de la máquina. De nuevo, el estado inicial se identifica por una doble flecha que apunta hacia él. La salida asociada con un estado se coloca bajo el nombre del estado, separada de éste por una barra horizontal. La transición desde un estado a otro se indica mediante una arista dirigida y etiquetada con el(los) correspondiente(s) carácter(es) de entrada. Efectivamente, las figuras 7.5 y 7.6 describen la misma máquina de estado finito, como puede verificar rápidamente el lector.

7.3 MÁQUINAS DE ESTADO FINITO COMO MODELOS DE SISTEMAS FÍSICOS

Una máquina de estado finito puede ser útil para modelar un sistema físico. En la sección 7.1 vimos cómo una máquina vendedora puede ser modelada como una máquina de estado finito. Mostraremos más ejemplos en esta sección.

Consideremos el problema de diseñar un contador módulo 3 que reciba una sucesión de números 0,1 y 2 como entrada, y produzca una sucesión de números 0,1 y 2 como salida, de manera que en cualquier instante la salida sea igual al módulo 3 de la suma de los dígitos de la sucesión de entrada[†] La máquina de la figura 7.7 generará la sucesión de salida como

		Entrada		
Estado	0	1	2	Salida
⇒ *A*	*A*	*B*	*C*	0
B	*B*	*C*	*A*	1
C	*C*	*A*	*B*	2

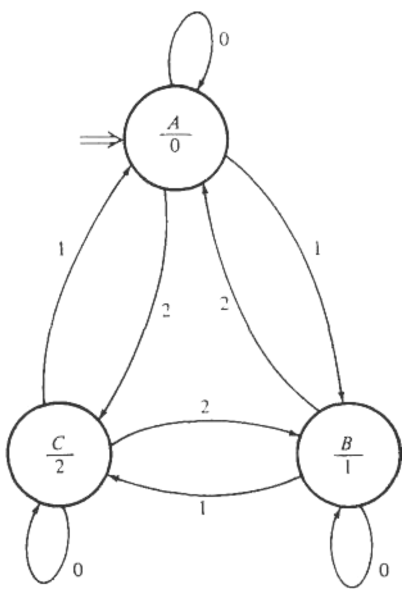


Figura 7.7

† El módulo 3 de la suma de un conjunto de números enteros es el residuo de la suma de los enteros cuando ésta es dividida por 3.

	Estado	Entrada 00	01	10	11	Salida
⇒	A	A	C	B	A	IGUAL
	B	B	C	B	B	MAYOR
	C	C	C	B	C	MENOR

Figura 7.8

	Estado	Entrada TAREA	FIESTA	MAL EXAMEN	Salida
⇒	A (FELIZ)	A	A	B	CANTAR
	B (ENOJADO)	C	A	B	MALDECIR
	C (DEPRIMIDO)	C	A	C	DORMIR

Figura 7.9

se especifica. Observe que A es un estado correspondiente a la situación en que el módulo 3 de la suma de todos los dígitos de entrada es 0; B es un estado que corresponde a la situación en que el módulo 3 de la suma de todos los dígitos de entrada es 1, y C es un estado que corresponde a la situación en que el módulo 3 de la suma de todos los dígitos de entrada es 2.

Ahora analicemos otro ejemplo, en el cual se diseña un dispositivo que compara dos números binarios para determinar si éstos son iguales, o cuál de los dos es mayor. Suponemos que los dígitos de los dos números llegan uno por uno, empezándose con los dígitos de orden inferior. Así, el alfabeto de entrada es {00,01, 10,11}, donde los dos dígitos de cada pareja son los dígitos correspondientes en los números a ser comparados. El alfabeto de salida es *{IGUAL, MAYOR, MENOR}*. La figura 7.8 muestra la máquina.

Por último, la figura 7.9 muestra un ejemplo de una máquina de estado finito que modela el comportamiento de un estudiante. El conjunto de estados es *{FELIZ, ENOJADO, DEPRIMIDO}*, el conjunto de entradas es *{TAREA, FIESTA, MAL EXAMEN}*, y el conjunto de salidas es *{CANTAR, MALDECIR, DORMIR}*. Aunque éste es un modelo muy simplificado, de hecho, abarca un cierto aspecto de cómo reacciona un estudiante bajo diversas condiciones (debemos señalar que éste no es un ejemplo vano incluido aquí para impresionar al lector. En realidad, los psicólogos, sociólogos, economistas y científicos de diversas disciplinas hacen uso de máquinas de estado finito para modelar los sistemas que ellos estudian).

7.4 MÁQUINAS EQUIVALENTES

Diremos que dos máquinas de estado finito son *equivalentes* si, a partir de sus respectivos estados iniciales, producen la misma sucesión de salida cuando se les da la misma sucesión de entrada. En otras palabras, máquinas equivalentes poseen comportamientos terminales idénticos aunque sus estructuras internas pueden ser diferentes. Por ejemplo, las dos máquinas de las figuras 7.10a y 7.106 son equivalentes. Exhortamos al lector a confirmar que, por ejemplo, para la sucesión de entrada 1122212212 ambas máquinas producen la

Estado	Entrada 1	Entrada 2	Salida
⇒ A	B	C	0
B	F	D	0
C	G	E	0
D	H	B	0
E	B	F	1
F	D	H	0
G	E	B	0
H	B	C	1

a)

Estado	Entrada 1	Entrada 2	Salida
⇒ A	B	C	0
B	C	D	0
C	D	E	0
D	E	B	0
E	B	C	1

b)

Figura 7.10

sucesión de salida 00010100001.[†] Es deseable que seamos capaces de determinar cuándo dos máquinas de estado finito son equivalentes, o dada una máquina de estado finito cómo encontrar una máquina equivalente que tenga menos estados, si es posible.[‡]

Para identificar máquinas equivalentes, introducimos la noción de *estados equivalentes.* En una máquina de estado finito, diremos que los estados s_i y s_j son *equivalentes* si para cualquier sucesión de entrada la máquina producirá la misma sucesión de salida, independientemente de que ésta inicie en s_i o en s_j. Es obvio que dos estados equivalentes pueden combinarse en uno solo sin cambiar el comportamiento terminal de la máquina. En específico, si s_i y s_j son estados equivalentes, podemos modificar la función de transición al eliminar el estado s_j y al remplazar todas las transiciones hacia el estado s_j con otras hacia s_i. Por ejemplo, para la máquina de estado finito de la figura 7.10a, puesto que son estados equivalentes C y F, D y G y E y H, podemos eliminar F, G y H para obtener la máquina de la figura 7.106.

Debemos recordar que el estado de una máquina representa el resumen de la historia de la máquina. Así, dos estados son equivalentes si representan resúmenes equivalentes en lo que al comportamiento terminal de la máquina se refiere. Por ejemplo, la máquina de la figura 7.10a acepta una sucesión de números 1 y 2 como entradas, y producirá un 1 si la suma de todos los dígitos que ha recibido es divisible por 4. En efecto, los estados E y H representan que la suma de los dígitos que la máquina ha recibido sea un múltiplo de 4, los estados C y F representan que la suma de los dígitos que la máquina ha recibido sea un múltiplo de 4 más 2, y los estados D y G representan que la suma de los dígitos que la máquina ha recibido sea un múltiplo de 4 más 3 (el estado B representa que la suma de dígitos que la máquina ha recibido sea un múltiplo de 4 más 1). En consecuencia, podemos combinar estados equivalentes sin cambiar el comportamiento terminal de la máquina.

---

† De acuerdo con nuestro modelo de la sección 7.2, una máquina de estado finito produce un símbolo de salida (aquel del estado inicial), antes de la llegada del primer símbolo de entrada. En consecuencia, siempre habrá un símbolo más en la sucesión de salida que en la correspondiente sucesión de entrada.

‡ Aunque aquí no abarcamos el punto de la construcción de máquinas de estado finito mediante dispositivos electrónicos, intuitivamente es claro que será menos costoso construir máquinas de estado finito con menores estados.

Estado	Entrada		Salida
	0	1	
A	B	F	0
B	A	F	0
C	G	A	0
D	H	B	0
E	A	G	0
F	H	C	1
G	A	D	1
H	A	C	1

Figura 7.11

Nuestra pregunta inmediata es cómo determinar cuándo dos estados son equivalentes; de acuerdo con la definición, para comprobar si dos estados son equivalentes es necesario un examen exhaustivo de todas las posibles sucesiones de entrada de longitud arbitraria. Ahora mostraremos un procedimiento efectivo para alcanzar ese propósito. Diremos que dos estados son *O-equivalentes* si tienen la misma salida; dos estados son 1 *-equivalentes* si tienen la misma salida y si, para todo carácter de entrada, sus sucesores son O-equivalentes. En general, diremos que dos estados son *k-equivalentes* si tienen la misma salida y si, para cada carácter de entrada, sus sucesores son (k - l)-equivalentes. Por ejemplo, para la máquina de estado finito de la figura 7.11, los estados A y C son O-equivalentes, y los estados G *y* H son 1-equivalentes. De la definición se obtiene directamente que si dos estados s_i y s_j son k-equivalentes, entonces para cualquier sucesión de entrada de longitud k o menor lamáquina producirá sucesiones de salida idénticas sin importar si está en el estado s_i o en el s_j. Es evidente que dos estados son equivalentes si son ¿-equivalentes para toda k.

Observamos que si s_i y s_j son k-equivalentes, y que si s_i y s_h son k-equivalentes, entonces s_j y s_h también son k-equivalentes; podemos definir una *relación de equivalencia* sobre el conjunto de todos los estados de modo que dos estados están relacionados si son ¿-equivalentes. Por tanto, esta relación induce una partición sobre el conjunto de todos los estados. Denotaremos esta partición mediante π_k. Por ejemplo, para la máquina de estado finito de la figura 7.11, notamos que

$$\pi_0 = \{\overline{ABCDE}\ \ \overline{FGH}\}$$
$$\pi_1 = \{\overline{ABE}\ \ \overline{CD}\ \overline{F}\ \ \overline{GH}\}$$
$$\pi_2 = \{\overline{AB}\ \ \overline{CD}\ \overline{E}\ \ \overline{F}\ \ \overline{GH}\}$$

Mostraremos cómo calcular estas particiones en el siguiente teorema.

Teorema 7.1 Dos estados están en el mismo bloque π_k si y sólo si están en el mismo bloque π_{k-1} y, para todo carácter de entrada, sus sucesores están en el mismo bloque π_{k-1}.

DEMOSTRACIÓN Sean s_i y s_j dos estados k-equivalentes. De acuerdo con la definición tienen la misma salida y, para todo carácter de entrada, sus sucesores son (¿ -l)-equivalentes. Observemos que de acuerdo con la definición, s_i y s_j también son (*k* - l)-equivalentes. Así, concluimos que los dos estados son *k*-equivalentes si y sólo si son (*k* - l)-equivalentes, y para todo carácter de entrada, sus sucesores también son (*k* - l)-equivalentes. Por tanto, el teorema se cumple.

El teorema 7.1 sugiere un procedimiento para calcular las particiones $\pi_0, \pi_1, \pi_2, \ldots,$ π_k en forma sucesiva. Por ejemplo, para la máquina de estado finito de la figura 7.11, primero notamos que

$$\pi_0 = \{\overline{ABCDE}\ \ \overline{FGH}\}$$

debido a que *A, B, C, D* y *E* tienen la misma salida, y *F, G* y *H* tienen la misma salida. Observamos que *A* y *B* están en el mismo bloque π_0 y, además, para la entrada 0 sus sucesores son *B* y *A*, los cuales están en el mismo bloque π_0, y para la entrada 1 sus sucesores son *F*, los cuales, trivialmente, están en el mismo bloque de π_0. Así, podemos concluir que *A* y *B* están en el mismo bloque π_1. De modo similar, debido a que *A* y *E* están en el mismo bloque 7i$_0$, para la entrada 0 sus sucesores *A* y *B* están en el mismo bloque π_0, y para la entrada 1 sus sucesores *F* y *G* están en el mismo bloque π_0, sabemos que *A* y *E* estarán en el mismo bloque π_1. Por otro lado, aun cuando *A* y *C* están en el mismo bloque π_0, sus sucesores *B* y *G* para la entrada 0 no están en el mismo bloque π_0, lo que significa que *A y C* no estarán en el mismo bloque π_1. Obtenemos

$$\pi_1 = \{\overline{ABE}\ \ \overline{CD}\ \overline{F}\ \ \overline{GH}\}$$

De manera similar,

$$\pi_2 = \{\overline{AB}\ \ \overline{CD}\ \overline{E}\ \ \overline{F}\ \ \overline{GH}\}$$
$$\pi_3 = \{\overline{AB}\ \ \overline{CD}\ \overline{E}\ \ \overline{F}\ \ \overline{GH}\}$$

Observemos lo siguiente:

1. Si π_k es igual a π_{k-1}, entonces π_m es igual a π_{k-1} para todo $m \geqq k$ (porque π_{k+1} se construye a partir de π exactamente de la misma manera como π_k se construye a partir de π_{k-1}).
2. π_k es un refinamiento de π_{k-1} (debido a que no es posible que dos estados sean *k*-equivalentes a menos que sean $(k-1)$-equivalentes).

La observación 1 nos permite terminar el procedimiento de construcción siempre que alcancemos un punto en que dos particiones sucesivas sean idénticas. En ese caso, todos los estados que sean ¿-equivalentes son equivalentes. Además, la observación 2 garantiza que nuestro procedimiento de construcción no irá más allá de π_{n-2}, donde *n* es el número de estados en la máquina, dado que π_0. tiene al menos dos bloques y π_{n-2} tendrá *n* bloques si el procedimiento de construcción no se termina antes. Para el ejemplo de la figura 7.11, como π_2 es igual a π_3, podemos concluir que los estados *A* y *B* son equivalentes, *C y D* son equivalentes, y *G y H* son equivalentes.

Para una determinada máquina de estado finito, podemos emplear el procedimiento presentado con anterioridad para determinar los estados equivalentes, y así obtener una máquina equivalente que podría tener menos estados. Por último, una vez que sabemos cómo determinar estados equivalentes en una máquina, estamos capacitados para determinar cuándo dos máquinas dadas son equivalentes. Exhortamos al lector a revisar el problema 7.9.

7.5 MÁQUINAS DE ESTADO FINITO COMO RECONOCEDORES DE LENGUAJE

En la sección 7.3 analizamos varios ejemplos de cómo modelar sistemas físicos mediante máquinas de estado finito. Ahora veremos que las máquinas de estado finito también pueden usarse naturalmente como instrumentos para reconocer (aceptar) enunciados de un lenguaje. Sea $O = \{0,1\}$ el alfabeto de salida de una máquina de estado finito. Diremos que un estado es de *aceptación* si su salida es 1; y un estado es de *rechazo* si su salida es 0. Por tanto, una sucesión de entrada es *aceptada* por una máquina de estado finito si tal sucesión conduce a la máquina desde el estado inicial hasta un estado de aceptación. Por otro lado, diremos que una sucesión de entrada es *rechazada* por una máquina de estado finito, si la sucesión conduce a la máquina desde el estado inicial hasta un estado de rechazo. Analicemos los siguientes ejemplos.

Ejemplo 7.1

La figura 7.12 muestra una máquina de estado finito que acepta todas las sucesiones binarias que terminan con los dígitos 011 (cuando una máquina de estado finito es utilizada como un aceptor, los estados de la máquina se dividen en sólo dos clases, a saber, estados de aceptación y de rechazo. Por tanto, introducimos la notación ligeramente más simple, de inscribir en circunferencias los nombres de los estados de aceptación -en lugar de escribir la salida de cada estado- como en la figura 7.12, donde D es el único estado de aceptación).

Ejemplo 7.2

La figura 7.13 muestra una máquina de estado finito que acepta todas las sucesiones binarias que tienen la forma de cualquier cantidad de números 0, seguido por un uno o más números 1, seguido por uno o más números 0, seguidos por un 1, seguido por cualquier cantidad de números 0, seguidos por un 1, y entonces seguidos por cualquier cosa.

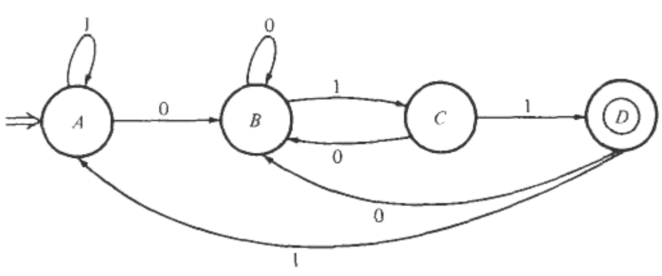


Figura 7.12

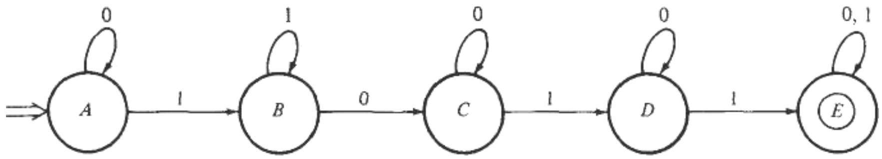


Figura 7.13

Diremos que un lenguaje es un *lenguaje de estado finito* si existe una máquina de estado finito que acepte exactamente todos los enunciados del lenguaje. Así, de acuerdo con el ejemplo 7.1, el lenguaje consistente en todas las sucesiones binarias que terminan con 011 es un lenguaje de estado finito. El ejemplo 7.2 muestra otro lenguaje de estado finito. Es claro que cualquier máquina de estado finito define un lenguaje de estado finito. Por otro lado, un lenguaje dado podría o no podría ser un lenguaje de estado finito. Ahora mostraremos que existen lenguajes que no son lenguajes de estado finito.† En la sección 7.6, mostraremos la forma de construir una máquina de estado finito para aceptar un lenguaje dado, el cual se sabe que es un lenguaje de estado finito.

Ejemplo 7.3

Demostremos que el lenguaje

$$L = \{a^k b^k \mid k \geqq 1\}$$

no es un lenguaje de estado finito. Supongamos lo contrario -existe una máquina de estado finito que acepta los enunciados de *L*. Esta máquina tiene Agestados. Es obvio que la máquina acepta el enunciado $a^N b^N$. A partir del estado inicial, la máquina visitará *N* estados después de recibir las *N* letras a en la sucesión de entrada como se ilustra en la figura 7.14a, donde s. es el estado inicial y $s_{j_1}, s_{j_2}, \ldots, s_{j_N}$ son los estados en que la máquina estuvo después de haber recibido la sucesión a^N. También, $s_{j_{2N}}$ es el estado en que la máquina se encuentra después de haber recibido la sucesión $a^N b^N$. Es obvio que, $s_{j_{2N}}$ es un estado de aceptación. De acuerdo con el principio del "palomar", de entre los $N + 1$ estados $s_{j_0}, s_{j_1}, s_{j_2}, \ldots, s_{j_N}$, hay dos de ellos que son el mismo. Supongamos que la maquina alcanza el estado s_k dos veces como se muestra en la figura 7.14o, y hay

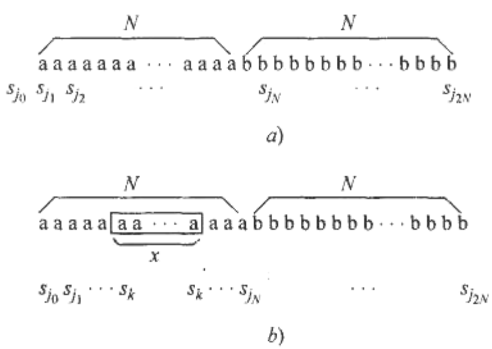


Figura 7.14

† De nuevo, tenemos otro ejemplo de cómo demostrar que alguna tarea (en este caso, encontrar una máquina de estado finito que acepte un lenguaje dado) es imposible.

x letras a entre la primera y segunda visita al estado $s_{i_{x}}$. Entonces la sucesión,

$$a^{N-x}b^{N}$$

la cual no es un enunciado del lenguaje, también será aceptada por la máquina de estado finito. En consecuencia, podemos concluir que el lenguaje L no es un lenguaje de estado finito.

Ejemplo 7.4 Demostremos que el lenguaje

$$L = \{a^k \mid k = i^2, i \geqq 1\}$$

no es un lenguaje de estado finito. Supongamos de nuevo que existe una máquina de estado finito que acepta al lenguaje L. Denotemos como N el número de estados de la máquina. Sea i un entero lo suficientemente grande de manera que

$$(i+1)^2 - i^2 > N$$

Consideremos la situación ilustrada de la figura 7.15. Puesto que entre la i^2-ésima a y la $(i+1)^2$-ésima a, la máquina de estado finito visitará un cierto estado s_k más de una vez, la eliminación de las letras a entre estas dos visitas dará lugar a una sucesión que también será aceptada por la máquina de estado finito. No obstante, esta sucesión no es un enunciado en el lenguaje debido a que contiene más de i^2 pero menos de $(i+1)^2$ letras a. Así, concluimos que L no es un lenguaje de estado finito.

Estos dos ejemplos ilustran un resultado general, conocido en la literatura como el *lema de bombeo* para lenguajes de estado finito.

Teorema 7.2 Sea L un lenguaje de estado finito aceptado por una máquina de estado finito con N estados. Para cualquier sucesión α en el lenguaje, cuya longitud es mayor o igual a N, α puede escribirse como uvw de manera que v es no-vacía y uv^iw también está en el lenguaje para $i \geqq 0$, donde v^i denota la concatenación de i copias de la sucesión v (en otras palabras, $uw, uvw, uvvw, uvvvw, \ldots$ están todas en el lenguaje).

DEMOSTRACIÓN Sin pérdida de generalidad, consideramos que la longitud de α es N. Sea $\alpha = a_1a_2a_3 \cdots a_N$. Sean $s_{j_0}, s_{j_1}, s_{j_2}, \ldots, s_{j_N}$ los estados que la máquina visita, donde s_{j_0} es el estado inicial y s_{j_N} es un estado de aceptación. Ahora de nuevo, de entre los $N + 1$ estados $s_{j_0}, s_{j_1}, s_{j_2}, \ldots, s_{j_N}$, existen dos de ellos que son el mismo. Supongamos que es el estado s_k, como se muestra de la figura 7.16. Si dividimos α en tres segmentos como se muestra de la figura 7.16, observamos que las sucesiones $uw, uvw, uvvw, uvvvw, \ldots, uv^iw, \ldots$ conducirán a la máquina desde el estado inicial s_{j_0} hasta el estado de aceptación s_{j_N}.

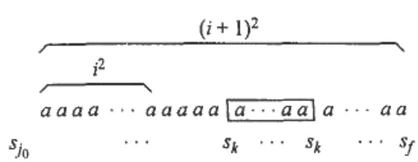


Figura 7.15

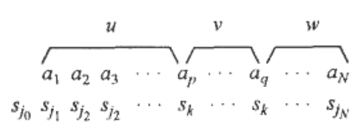


Figura 7.16

*7.6 LENGUAJES DE ESTADO FINITO Y LENGUAJES TIPO-3

Algo que sorprende, es la estrecha relación que hay entre los lenguajes de estado finito y los lenguajes tipo-3. En efecto, mostraremos que un lenguaje de estado finito es un lenguaje tipo-3 y que un lenguaje tipo-3 es un lenguaje de estado finito.[†] Para ello, introducimos la noción de máquina de estado finito *no-determinística.* Una máquina de este tipo difiere de una máquina de estado finito determinística, como se definió con anterioridad, en que la función de transición para una máquina no-determinística es una función desde $S \times I$ hasta $\mathscr{P}(S)$. Además, debido a que las máquinas de estado finito no-determinísticas tienen un uso exclusivo como reconocedora de lenguaje, los caracteres de salida se restringen por lo común a 0 y 1. Una máquina de estado finito no-determinística se especifica formalmente como sigue:

1. Un conjunto finito de estados $S = \{s_0, s_1, s_2, \ldots\}$.
2. Un elemento especial del conjunto S, s_0, conocido como el *estado inicial.*
3. Un conjunto finito de caracteres de entrada $I = \{i_1, i_2, \ldots\}$
4. Un conjunto finito de caracteres de salida $O = \{o_1, o_2, \ldots\}$.
5. Una función F desde $S \times I$ a $\mathscr{P}(S)$, conocida como la *función de transición.*
6. Una función g desde S hacia O, conocida como la *función de salida.*

Cuando una máquina de estado finito no-determinística se encuentra en cierto estado y recibe algún carácter de entrada, ésta puede tener más de un estado siguiente. En realidad, la única distinción entre una máquina de estado finito no-determinística y una determinística es que en la función de transición mapea un par ordenado de un estado y un carácter de entrada (el estado de la máquina y el carácter de entrada) hacia un subconjunto de estados (todos los posibles estados siguientes) en lugar de un solo estado. Podemos imaginar que la máquina entrará en todos estos posibles estados siguientes (una forma útil de visualizar esto es imaginar que la máquina hace varias copias de ella misma, y cada copia se encuentra en uno de los posibles estados siguientes). A partir de cada uno de los estados siguientes y al recibir otro carácter de entrada, la máquina entrará de nuevo en todos los posibles estados siguientes. La figura 7.17a muestra una máquina de estado finito no-determinística, donde una transición puede tener más de un estado siguiente. Por ejemplo, cuando la máquina inicia en el estado *A* y recibe una entrada de 1, la máquina ingresará en los estados *B* y C de modo simultáneo. La figura 7.17*b* muestra los estados de la máquina correspondientes a la sucesión de entrada 000100001.

† Permítasenos hacer aquí una pequeña observación: si el estado inicial de una máquina de estado finito es de aceptación; entonces, de acuerdo con la definición de aceptación, la *sucesión nula* (una sucesión que no contiene símbolos) es un enunciado en el lenguaje aceptado por la máquina de estado finito. No obstante, de acuerdo con nuestro análisis en el capítulo 2, ignoramos la posibilidad de tener a la sucesión nula como un enunciado en un lenguaje tipo-3. Seremos consistentes si ignoramos a la sucesión nula como una sucesión que conduce a una máquina de estado finito desde el estado inicial hasta un estado de aceptación cuando el estado inicial de la máquina es un estado de aceptación.

	Estado	Entrada 0	Entrada 1	Salida
⇒	A	B	B, C	0
	B	A, C	C	0
	C	A	B, C	1

a)

	0	0	0	1	0	0	0	0	1
A	B	A	A	B	A	A	A	A	B
		C	B	C	C	B	B	B	C
							C	C	

b)

Figura 7.17

Uno puede señalar de inmediato que una máquina de estado finito no-determinística no es un buen modelo para cualquier sistema físico, puesto que ningún sistema físico puede hacer una réplica de sí mismo para cada transición y de modo indefinido. No obstante, veremos que una máquina no-determinística es un modelo abstracto muy útil, y estamos particularmente interesados en utilizarla como un reconocedor de lenguaje. Diremos que una sucesión es aceptada por una máquina de estado finito no-determinística si al comenzar desde el estado inicial, de entre todos los estados finales a los que la sucesión conducirá a la máquina, uno de ellos es un estado de aceptación. Por ejemplo, para la máquina de estado finito no-determinística de la figura 7.17*a*, la sucesión 0001 es aceptada, en tanto que la sucesión 11100 no lo es.

Uno podría preguntarse naturalmente, en lo que a un reconocedor de lenguaje se refiere, si en verdad una máquina de estado finito no-determinística es más poderosa que una determinística, en el sentido de que existen lenguajes que pueden ser reconocidos por una no-determinística, pero no pueden ser reconocidos por una determinística (debido a que el modelo de una máquina no-determinística incluye el modelo de una máquina determinística como un caso especial, una máquina no-determinística es al menos tan poderosa como una determinística). Sin embargo, resulta que las máquinas no-determinísticas no son más poderosas que las máquinas determinísticas. Ahora mostraremos que para cualquier máquina de estado finito no-determinística, existe una máquina de estado finito determinística que acepta exactamente el mismo lenguaje. Sea Muña máquina de estado finito no-determinística con:

1. $\{s_0, s_1, s_2, \ldots\}$ como el conjunto de estados.
2. s_0 es el estado inicial.
3. $\{i_1, i_2, \ldots\}$ es el conjunto de caracteres de entrada.
4. $\{0, 1\}$ es el conjunto de caracteres de salida.
5. F de $S \times I$ a $\mathscr{P}(S)$ es la función de transición.
6. g de S a $\{0, 1\}$ es la función de salida.

Estado	Entrada 0	Entrada 1	Salida
⇒ A	B	B, C	0
B	A, C	—	0
C	A	B, C	1

a)

Estado	Entrada 0	Entrada 1	Salida
⇒ $\{A\}$	$\{B\}$	$\{B, C\}$	0
$\{B\}$	$\{A, C\}$	$\{\ \}$	0
$\{C\}$	$\{A\}$	$\{B, C\}$	1
$\{A, B\}$	$\{A, B, C\}$	$\{B, C\}$	0
$\{A, C\}$	$\{A, B\}$	$\{B, C\}$	1
$\{B, C\}$	$\{A, C\}$	$\{B, C\}$	1
$\{A, B, C\}$	$\{A, B, C\}$	$\{B, C\}$	1
$\{\ \}$	$\{\ \}$	$\{\ \}$	0

b)

Figura 7.18

Construimos la correspondiente máquina de estado finito determinística $\hat{M}$ como sigue. Sean

1. $\mathscr{P}(S)$ el conjunto de estados.
2. $\{s_0\}$ el estado inicial.
3. $\{i_1, i_2, \ldots\}$ el conjunto de caracteres de entrada.
4. $\{0, 1\}$ el conjunto de caracteres de salida.
5. $\hat{F}$ de $\mathscr{P}(S) \times I$ a $\mathscr{P}(S)$ la función de transición tal que $\hat{F}(\{\ \}, i_a) = \{\ \}$ para cualquier carácter de entrada i_q y para cualquier subconjunto no vacío S_p, de S, $\hat{F}(S_p, i_q) = \bigcup_{s_r \in s_p} F(s_r, i_q)$.
6. $\hat{g}$ de $\mathscr{P}(S)$ a $\{0, 1\}$ la función de salida tal que para cualquier $S_p \subseteq S$, $\hat{g}(S_p)$ es igual a 1 si existe un s_r en S_p tal que $g(s_r)$ sea igual a 1, y sea igual a 0 en otro caso.

La figura 7.18 muestra un ejemplo. Para la máquina de estado finito no-determinística -mostrada en la figura 7.18*a*- la máquina de estado finito determinística correspondiente i$\hat{M}$ se muestra en la figura 7.18*b* (en la figura 7.18*a* seguimos la convención en la de usar un "—" en lugar del símbolo de conjunto vacío). Una vez que el lector entienda la construcción, se convencerá de que las máquinas M y $\hat{M}$ aceptan exactamente el mismo lenguaje, puesto que una sucesión de entrada aceptada por *M* conducirá a $\hat{M}$ a un estado, el cual es un sub-conjunto de estados de *M*, en el que existe al menos uno que es un estado de aceptación en *M*. Por tanto, aquí no incluiremos la demostración formal.

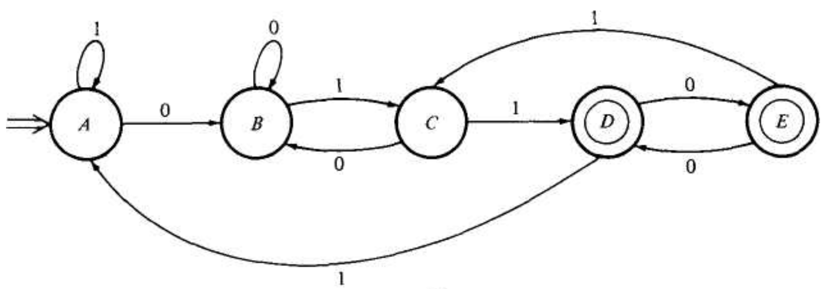
Ahora mostramos que la clase de lenguajes de estado finito es exactamente la clase de lenguajes de tipo-3; también, que dada una máquina de estado finito, podemos tener una gramática tipo-3 que especifique el lenguaje aceptado por la máquina de estado finito. Tambien demostraremos, que a partir una gramática tipo-3, podemos construir una máquina de estado finito no-determinística que acepte el lenguaje especificado por la gramática. Consideremos los siguientes ejemplos.

Ejemplo 7.5

Para la máquina de estado finito mostrada de la figura 7.19*a*, construimos una gramática que especifica el lenguaje aceptado por la máquina de la siguiente manera.

1. Sea $\{0, 1\}$, el conjunto de caracteres de entrada, el conjunto de terminales.
2. Sea $\{A, B, C, D, E\}$, el conjunto de estados, el conjunto de no-termmales.
3. Sea A, el estado inicial, el símbolo de inicio.
4. Correspondiente a la transición $f(s_p, i_q) = s_k$, existe una producción $s_p \to i_q s_k$ si s_k no es un estado de aceptación, y existen dos producciones $s_p \to i_q s_k$ y $s_p \to i_q$ si s_k es un estado de aceptación. Por ejemplo, correspondiente a la transición $f(A, 0) = B$, tenemos la producción $A \to 0B$, y correspondiente a la transición $f(C, 1) = D$, tenemos las producciones $C \to 1D$ y $C \to 1$.

Al obtener la gramática mostrada de la figura 7.19¿, observamos que cada no-terminal en la gramática representa un conjunto de sucesiones que conducen a la máquina de estado finito, desde el estado correspondiente hasta un estado de aceptación. Uno podría estar de acuerdo con que la gramática de la figura 7.19*b* especifica el lenguaje aceptado por la máquina de la figura 7.19*a*.



a)

$A \to 0B$
$A \to 1A$
$B \to 0B$
$B \to 1C$
$C \to 0B$
$C \to 1D$
$C \to 1$
$D \to 0E$
$D \to 0$
$D \to 1A$
$E \to 0D$
$E \to 0$
$E \to 1C$

Figura 7.19

Ejemplo 7.6

Para la gramática de la figura 7.20a, construimos una máquina de estado finito que acepte los enunciados del lenguaje especificado por la gramática de la siguiente manera.

1. Sea {0, 1}, el conjunto de terminales, el conjunto de caracteres de entrada.
2. Existe un estado que corresponde a cada no-terminal, con el estado correspondiente al símbolo de inicio como el estado inicial (el estado *A* en este caso). Existe un estado adicional *E,* el cual es un estado de aceptación. Además, existe otro estado, *T,* que es llamado un estado de *atrapamiento.* Siempre que una máquina entra en un estado de atrapamiento, permanecerá en dicho estado y no podrá pasar a ningún otro estado. Así, si una sucesión de entrada conduce a la máquina de estado finito hacia un estado de atrapamiento, posiblemente tal sucesión de entrada no puede ser una porción de un enunciado en el lenguaje.
3. Para una producción de la forma $N_p \rightarrow i_q N_k$, existe una transición del estado N_p hasta el estado *Nk* cuando la entrada es i_q. Por ejemplo, en correspondencia con la

$A \rightarrow 0A$
$A \rightarrow 1B$
$B \rightarrow 0C$
$B \rightarrow 0D$
$C \rightarrow 0$
$C \rightarrow 1B$
$C \rightarrow 1D$
$D \rightarrow 1$
$D \rightarrow 1A$

a)

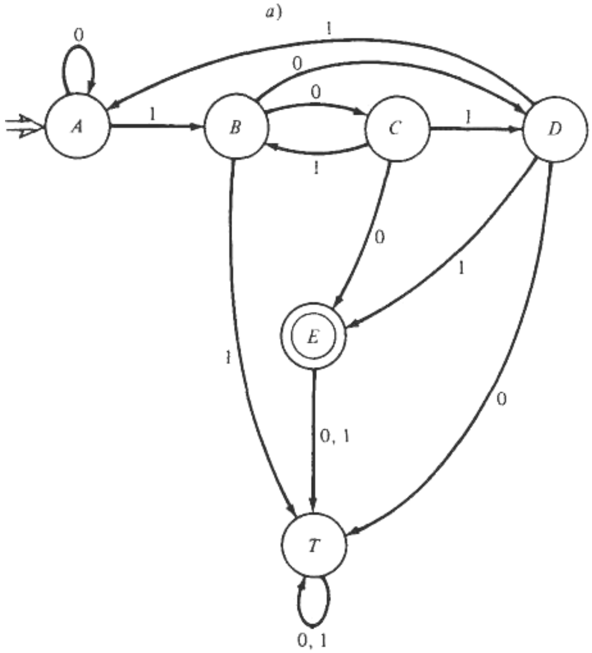


Figura 7.20

b)

Producción $A \rightarrow 1B$ existe una transición desde el estado A hasta el estado B, cuando el carácter de entrada es 1. Para una producción de la forma $N_p \rightarrow i_q$, existe una transición del estado N_p al estado de aceptación E cuando la entrada es i_q. Por ejemplo, en correspondencia con la producción $C \rightarrow 0$, existe una transición del estado C al estado E cuando el carácter de entrada es 0. Por otro lado, si no hay una producción de la forma $N_p \rightarrow i_q N_k$ o $N_p \rightarrow i_q$ para el estado N_p y la entrada i_q, existe una transición desde N_p hasta el estado de atrapamiento T cuando la entrada es i_q. Por ejemplo, existe una transición desde el estado B hasta el estado T cuando el carácter de entrada es 1. Por último, para cualquier carácter de entrada, existe una transición desde el estado de aceptación E hasta el estado de atrapamiento T.

Así, obtenemos la máquina de estado finito no-determinística mostrada en la figura *7.20b*. De nuevo, confiamos en que la construcción se representa con la suficiente claridad para convencer al lector de que la máquina de estado finito no-determinística que hemos construido acepta exactamente los enunciados del lenguaje especificado por la gramática de la figura 7.20a. Puesto que con anterioridad hemos demostrado que cualquier lenguaje que puede ser aceptado por una máquina de estado finito no-determinística, también puede ser aceptado por una determinística, el lenguaje especificado por la gramática de la figura 7.20a es en verdad un lenguaje de estado finito.

□

Esperamos que los ejemplos 7.5 y 7.6 tengan la suficiente claridad para convencer al lector de nuestra aseveración: la clase de lenguajes de estado finito es exactamente la clase de lenguajes tipo-3. Por tanto, no escribiremos aquí una demostración general.

7.7 NOTAS Y REFERENCIAS

Existen varias razones para la introducción del tópico de máquinas de estado finito: (1) para representar cómo pueden ser descritos los sistemas físicos mediante un modelo abstracto, (2) para ver la forma en que pueden ser descritos los dispositivos de reconocimiento de lenguajes mediante un modelo abstracto (un tópico de suma importancia en la construcción de un compilador), y (3) para observar la estrecha y fina conexión entre máquinas y lenguajes. De hecho, existe una jerarquía de máquinas -a saber, máquinas de estado finito, autómatas *pushdown,* autómatas linealmente acotadas y máquinas de Turing. También existe una jerarquía de lenguajes correspondiente -a saber, los lenguajes tipo 3, 2, 1 y 0, como fueron brevemente descritos en el capítulo 2. Las máquinas de estado finito pueden usarse como reconocedores para lenguajes tipo-3, las autómatas *pushdown* como reconocedores para lenguajes tipo-2, las autómatas linealmente acotadas como reconocedores para lenguaje tipo-1, y las máquinas de Turing como reconocedores de lenguaje tipo-O. Como referencias sobre el tópico de máquinas de estado finito, véase Gilí [1], Hennie [3] y Kohavi [5]. Como referencias generales sobre máquinas y lenguajes formales, véase Harrison [2], Hopcroft y Ullman [4] y Lewis y Papadimitriou [6].

1. Gill, A.: *Introduction to the Theory of Finite-State Machines,* McGraw-Hill Book Company, Nueva York, 1962.

2. Harrison, M. A.: *Introduction to Formal Language Theory,* Addison-Wesley Publishing Company, Reading, Mass., 1978.
3. Hennie, F. C: *Finite-State Models for Logical Machines,* John Wiley & Sons, Nueva York, 1968.
4. Hopcroft, J. E. y J. D. Ullman: *Introduction to Automata Theory, Languages and Computation,* Addison-Wesley Publishing Company, Reading, Mass., 1979.
5. Kohavi, Z.: *Switching and Automata Theory,* 2ª ed., McGraw-Hill Book Company, Nueva York, 1978.
6. Lewis, H. R. y C. H. Papadimitriou: *Elements of the Theory of Computation,* Prentice-Hall, Englewood Cliffs, N.J., 1981.

PROBLEMAS

7.1 Diseñe una máquina de estado finito con {0, 1} como alfabeto de entrada y {0, 1, 2} como su alfabeto de salida, de modo que para cualquier sucesión de entrada la correspondiente sucesión de salida consista en dos números 2 seguidos por la sucesión de entrada retrasada por una unidad de tiempo. Por ejemplo, podríamos tener:

Sucesión de entrada: 10001011001
Sucesión de salida: 22100010110

Observe que el primer símbolo de salida es la salida del estado inicial que ocurre antes del primer símbolo de entrada.

7.2 Diseñe una máquina de estado finito con {0, 1} como alfabeto de entrada y de salida, tal que la salida 1 será producida al empezar con el tercer 1 en cualquier bloque de tres o más números 1 en la sucesión de entrada.
Por ejemplo, podríamos tener:

Sucesión de entrada: 0011110101100111111010
Sucesión de salida: 0000011000000000111000

7.3 Una máquina de estado finito de tres estados tiene {0, 1} como sus alfabetos de entrada y salida. A partir de la siguiente sucesión de entrada y su correspondiente sucesión de salida, determine la máquina.

Sucesión de entrada: 00010101
Sucesión de salida: 011001110

7.4 Para la máquina mostrada en la figura 7P.1, la aplicación del símbolo de entrada 0 requiere 1 unidad de energía y la aplicación del símbolo de entrada 1 requiere 2 unidades de energía. Determine una sucesión de entrada de energía mínima que lleve a la máquina desde el estado *A* al estado *H.*

Estado	Entrada	
	0	1
A	*B*	C
B	*E*	*D*
C	*B*	*C*
D	*E*	*F*
E	*E*	*C*
F	*H*	*G*
G	*E*	*D*
H	*G*	*G*

Figura 7P.1

7.5 Una variación del modelo de una máquina de estado finito es el siguiente:

Un conjunto finito de estados $S = \{s_0, s_1, s_2, \ldots\}$.
Un elemento especial del conjunto S, s_0, es conocido como el *estado inicial.*
Un conjunto finito de caracteres de entrada $I = \{i_1, i_2, \ldots\}$ Un conjunto finito de caracteres de salida $O = \{o_1, o_2, \ldots\}$
Una función f de $S \times I$ a S, conocida como la *función de transición.*
Una función g de $S \times I$ a O, _conocida como la *función de salida.*

En este modelo la salida se asocia con cada transición de estado, en lugar de estarlo con cada estado, como en el modelo introducido en la sección 7.2.

a) Para la máquina de la figura 7P.2, las transiciones de estado y la salida asociada son separadas por una diagonal (/), determine la sucesión de salida correspondiente a la sucesión de entrada 110101.

b) Determine la sucesión de entrada que producirá la sucesión de salida 110101.

c) Muestre un procedimiento general para determinar una sucesión de entrada que produzca una sucesión de salida dada.

		Entrada	
	Estado	0	1
⇒	*A*	*B*/1	*C*/0
	B	*B*/1	*C*/0
	C	*D*/1	*C*/1
	D	*B*/0	*D*/0

Figura 7P.2

7.6 Para el modelo de máquina de estado finito del problema 7.5, diseñe una máquina de estado finito con $\{0, 1\}$ como sus alfabetos de entrada y salida, que produzca una sucesión de salida en que los símbolos de salida 3º, 6º, . . . , $3k^{\text{ésimo}}$ sean complemento de los símbolos de entrada correspondientes, mientras que los otros símbolos de salida sean idénticos a los correspondientes símbolos de entrada.

7.7 Sea $\{s_0, s_1, \ldots, s_n\}$ el conjunto de estados de una máquina de estado finito con s_0 como estado inicial. Sea $A^\star$ el conjunto de todas las sucesiones sobre el alfabeto de entrada^!. Definiremos una relación binaria R sobre $A^\star$ tal que para α_1 y α_2 en $A^\star$, $(\alpha_1, \alpha_2) \in R$ si y sólo si ambos α_1 y α_2 llevan la máquina de estado finito del estado s_0 a s_i para algún s_i. Demuestre que R es una relación de equivalencia.

7.8 Para la máquina de estado finito mostrada en la figura 7P.3:

a) Liste todos los estados 0-equivalentes.

b) Encuentre todos los estados equivalentes y obtenga una máquina de estado finito con el menor número de estados.

		Entrada		
	Estado	0	1	Salida
⇒	*A*	*F*	*B*	0
	B	*D*	*C*	0
	C	*G*	*B*	0
	D	*E*	*A*	1
	E	*D*	*A*	0
	F	*A*	*G*	1
	G	*C*	*H*	1
	H	*A*	*H*	1

Figura 7P.3

7.9 Demuestre que las dos máquinas de estado finito mostradas en la figura 7P.4 son equivalentes.

	Entrada		
Estado	0	1	Salida
⇒ A	B	C	0
B	B	D	0
C	A	E	0
D	B	E	0
E	F	E	0
F	A	D	1
G	B	C	1

a)

	Entrada		
Estado	0	1	Salida
⇒ A	H	C	0
B	G	B	0
C	A	B	0
D	D	C	0
E	H	B	0
F	D	E	1
G	H	C	1
H	A	E	0

b)

Figura 7P.4

7.10 Acerca de una partición sobre el conjunto de estados de una máquina de estado finito n, diremos que es una partición *preservada* si para cualesquiera dos estados s_i y s_k que están en el mismo bloque de n, entonces los dos estados $f(s_j, i)$ y $f(s_k, i)$ también están en el mismo bloque de n para cualquier símbolo de entrada i. Por ejemplo, para la máquina de estado finito mostrada en la figura 7P.5*a*, $\{\overline{AB}\ \overline{CD}\}$ es una partición preservada. Observe que tanto $\{\overline{ABCD}\}$ como $\{\overline{A}\ \overline{B}\ \overline{C}\ \overline{D}\}$ son particiones trivialmente preservadas.

a) ¿Puede encontrarse otra partición preservada sobre el conjunto de estados?

b) Para la máquina de estado finito mostrada en la figura 7P.5*b*, encuentre dos particiones preservadas π_1 y π_2 tales que

$$\pi_1 \cdot \pi_2 = \{\overline{A}\ \overline{B}\ \overline{C}\ \overline{D}\ \overline{E}\ \overline{F}\}$$

c) Demuestre que si π_1 y π_2 son dos particiones preservadas, también lo son $\pi_1 \cdot \pi_2$ y $\pi_1 + \pi_2$.

	Entrada	
Estado	0	1
A	B	C
B	A	D
C	D	A
D	C	B

a)

	Entrada	
Estado	0	1
A	F	B
B	B	A
C	C	E
D	E	F
E	D	C
F	A	D

b)

Figura 7P.5

7.11 Para cada uno de los conjuntos descritos enseguida, encuentre una máquina de estado finito determinística que reconozca los conjuntos:

a) $L = \{(01)^i 1^{2j} \mid i \geqq 1, j \geqq 1\}$

b) $L = \{0^i 1 0^j \mid i \geqq 1, j \geqq 1\}$

c) $L = \{0^i 1 0^j \mid i \geqq 1, j \geqq 1\} \cup \{0^k \mid k \geqq 3\}$

d) $L = \{\text{todas las cadenas binarias terminadas en } 001\} \cup \{1\}$

7.12 Para cada uno de los conjuntos descritos abajo, halle una máquina de estado finito determinística que reconozca los conjuntos:

a) El conjunto de cadenas de números 0 y 1 en cada una de las cuales hay una cantidad par de números 0.

b) El conjunto de cadenas de números 0 y 1 en cada una de las cuales la cantidad de números 1 no es múltiplo de 4.

c) El conjunto de cadenas de números 0 y 1 en cada una de las cuales la cantidad de números 0 es par y la cantidad de números 1 es un múltiplo de 3.

d) El conjunto de cadenas de números 0, 1 y 2 en cada una de las cuales cada 1 es seguido inmediatamente por al menos dos números 2 y cada 0 es precedido inmediatamente por al menos dos números 2.

e) El conjunto de cadenas de números 0 y 1 cada una de las cuales es una representación binaria de un entero de la forma $4k + 3$, para $k \geqq 1$.

f) El conjunto de cadenas de números 0 y 1 cada una de las cuales es una representación binaria de un entero de la forma $8^k + 1$, para $k \geqq 1$.

g) El conjunto de cadenas de números 0 y 1 cada una de las cuales inicia con un 1 y termina con 010.

h) El conjunto de cadenas de números 0 y 1 cada una de las cuales termina con 01^k, $k \geqq 1$.

i) El conjunto de cadenas de números 0 y 1 en las que cada 0 es seguido por tres o más números 1, con el 0 como primer símbolo.

j) El conjunto de cadenas de números 0 y 1 en las que cada bloque de números 1 contiene tres o más números 1 y es seguido por exactamente un 0, y los tres primeros o más símbolos son números 1.

k) El conjunto de cadenas de números 0 y 1, de las cuales ninguna contiene la subcadena 010.

7.13 Dé una descripción (verbal o en notación de conjuntos) de los conjuntos de cadenas reconocidos por cada una de las máquinas de estado finito en la figura 7P.6.

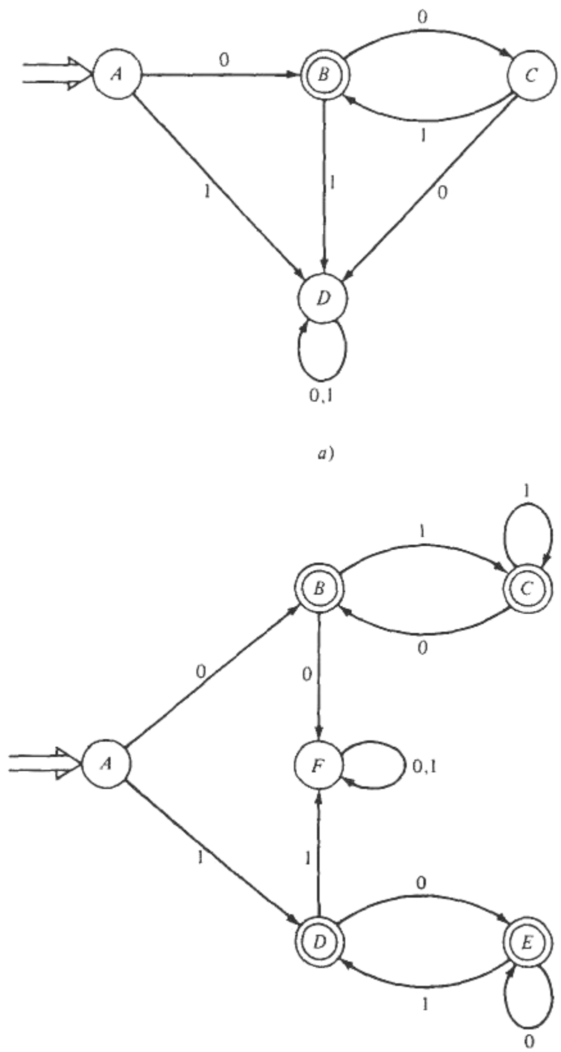
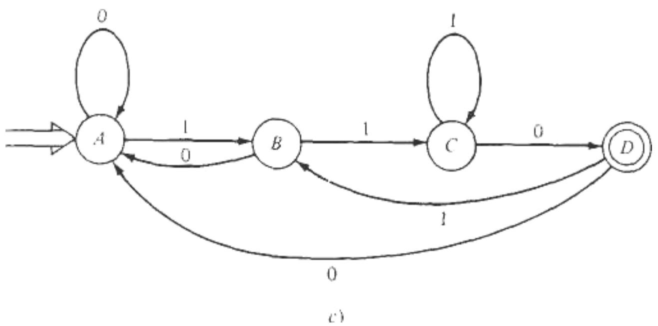


Figura 7P.6



c)

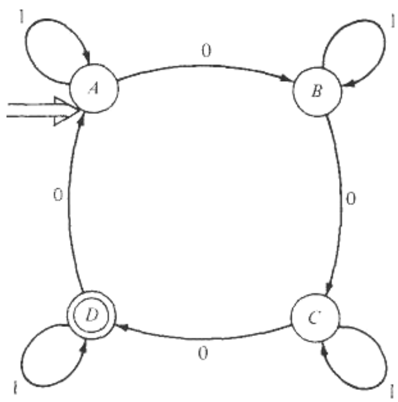


Figura 7P.6 (Cont.)

d)

7.14 Dada una máquina de estado finito determinística con su estado inicial no especificado, una sucesión de entrada es aceptada por la máquina si se puede escoger un estado inicial que permita a la sucesión de entrada llevar a la máquina desde el estado inicial escogido hasta un estado de aceptación.

a) ¿Aceptará la máquina de estado finito mostrada en la figura 7P.7 a la sucesión de entrada 0110 de acuerdo con esta nueva definición de aceptación?

b) Todos los lenguajes aceptados por máquinas de estado finito que usan esta definición de aceptación, ¿son lenguajes regulares? Demuestre su aseveración.

	Estado	Entrada 0	Entrada 1	Salida
⇒	*A*	*A*	*C*	0
	B	*C*	*A*	0
	C	*D*	*B*	0
	D	*B*	*D*	1

Figura 7P.7

c) Repita los incisos *a)* y *b)* si la definición de aceptación se cambia a la siguiente forma: una sucesión de entrada es aceptada por la máquina de estado finito si para cualquier elección arbitraria de un estado inicial, la sucesión de entrada llevará a la máquina desde el estado inicial elegido hasta un estado de aceptación.

7.15 *a)* El estado inicial de la máquina de estado finito mostrado en la figura 7P.8 es desconocido. Diseñe una sucesión de entrada que lleve a la máquina al estado *B* sin tomar en cuenta el estado inicial.

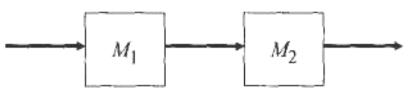
b) Una *sucesión sincronizada* es una sucesión de entrada que lleva a una máquina de estado finito a un estado final específico. Demuestre que si una máquina de estado finito con *n* estados tiene una sucesión sincronizada, entonces tiene una sucesión sincronizada de longitud menor o igual a $2^n - 2$.

c) Dé un ejemplo de una máquina de estado finito que no tenga una sucesión sincronizada.

	Entrada	
Estado	0	1
A	*B*	*A*
B	*A*	*D*
C	*C*	*A*
D	*B*	*C*

Figura 7P.8

7.16 Dos máquinas de estado finito con alfabetos de entrada y salida {0, 1}, pueden ser conectadas en *serie* como se muestra en la figura 7P.9*a*, donde el símbolo de salida de M, se utiliza como el símbolo de entrada de M_2. Para las máquinas M_1 y M_2 mostradas en la figura 7P.9*b*, determine una máquina de estado finito cuyo comportamiento terminal sea idéntico a la conexión en sucesión de M_1 y M_2.



a)

		Entrada		
	Estado	0	1	Salida
⇒	*A*	*A*	*C*	0
	B	*C*	*B*	0
	C	*B*	*A*	1

M_1

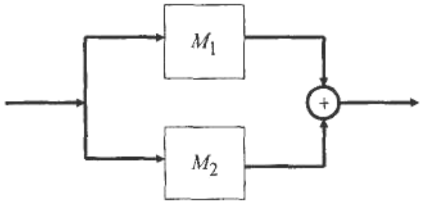
		Entrada		
	Estado	0	1	Salida
⇒	*D*	*D*	*E*	0
	E	*E*	*D*	1

M_2

b)

Figura 7P.9

7.17 Dos máquinas de estado finito con alfabetos de entrada y salida {0, 1} pueden ser conectadas en *paralelo* como se muestra en la figura 7P. 10*a*, donde cada símbolo de entrada es enviado simultáneamente a ambas máquinas, y la salida total es la "disyunción lógica" de los símbolos de salida de las dos máquinas. Para las máquinas M_1 y M_2 mostradas en la figura 7P.10*b*, determine una máquina de estado finito cuyo comportamiento terminal sea idéntico a la conexión en paralelo de M_l y M_2.



a)

	Entrada		
Estado	0	1	Salida
⇒ A	A	C	0
B	C	B	0
C	B	A	1

M_1

	Entrada		
Estado	0	1	Salida
⇒ D	D	E	0
E	E	D	1

M_2

b)

Figura 7P.10

7.18 Demuestre que si una máquina de estado finito acepta una sucesión de entrada "suficientemente larga", entonces también acepta un número infinito de sucesiones de entrada.
Sugerencia: ¿cómo definiría de manera precisa "suficientemente larga"?

7.19 Demuestre que cada uno de los siguientes lenguajes no es un lenguaje de estado finito.

a) $L = \{0^i 1^j \mid i \geqq j\}$

b) $L = \{0^i 1^j \mid i \leqq j\}$

c) $L = \{0^k \mid k = 2^i, i \geqq 1\}$

d) $L = \{1^i 0^j 1^{i+j} \mid i \geqq 1, j \geqq 1\}$

7.20 Demuestre que el lenguaje

$$L = \{xx \mid x \text{ es una cadena de ceros y unos}\}$$

no es un lenguaje de estado finito.

7.21 Construya una gramática para cada uno de los lenguajes aceptados por las máquinas de estado finito de los problemas 7.13 y 7.23.

7.22 Convierta cada una de las máquinas de estado finito no-determinísticas de la figura 7P. 11 a la forma determinística.

	Entrada		
Estado	0	1	Salida
⇒ A	B	A,C	0
B	C	A	1
C	A	–	0

	Entrada		
Estado	0	1	Salida
⇒ A	B,C	–	0
B	D	B	0
C	A	C	0
D	A	B,C	1

b)

	Entrada		
Estado	0	1	Salida
⇒ A	A,C	A,B	0
B	E	–	0
C	–	E	0
D	–	–	1
E	–	–	1

Figura 7P.11

c)

7.23 Dé una descripción (verbal o en notación de conjuntos) del conjunto de cadenas aceptadas por cada una de las máquinas de estado finito no-deterministicas de la figura 7P. 12 (si una transición conduce a un estado al conjunto vacío, la arista correspondiente es omitida).

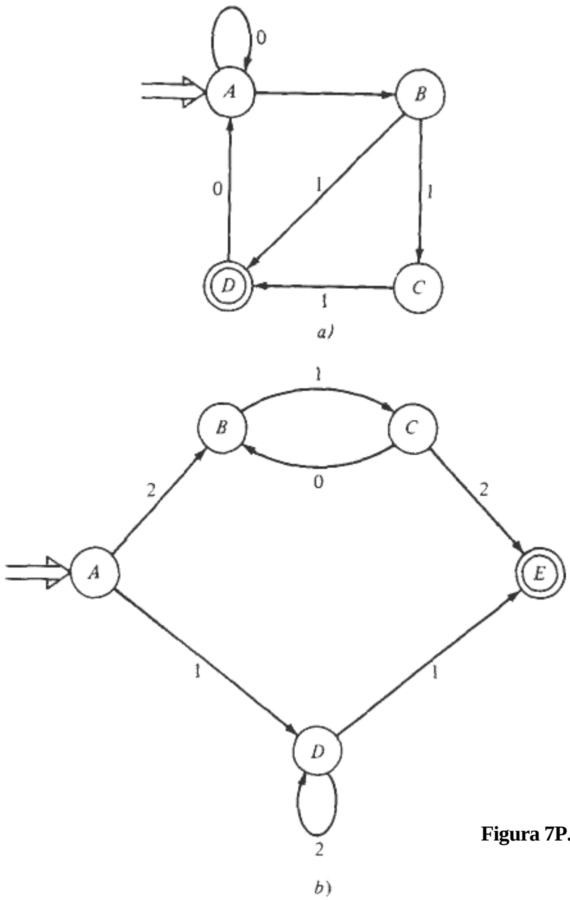


Figura 7P.12

7.24 En este problema, mostramos que la noción de las máquinas de estado finito no-determinísticas es muy útil para el diseño de máquinas determinísticas.

a) Diseñe una máquina de estado finito determinística con {0, 1} como su alfabeto de entrada que acepta todas las sucesiones terminadas tanto en 1010 como en 001.

b) ¿Cuál es el conjunto de sucesiones aceptadas por la máquina de estado finito no-determinística de la figura 7P.13? Convierta la máquina no-determinística en una determinística.

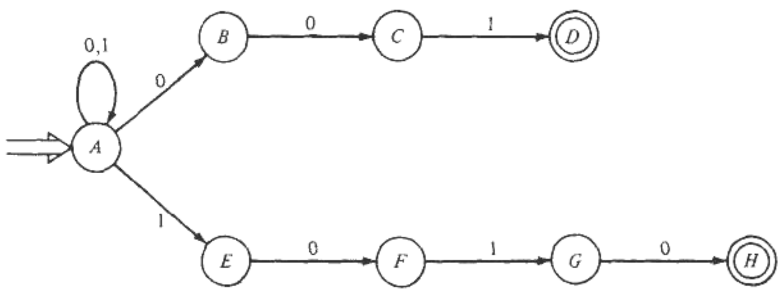


Figura 7P.13

7.25 Una sucesión que no contiene símbolos es referida como una *sucesión nula,* la cual comúnmente se denota por λ. La noción de una sucesión nula puede entenderse fácilmente si observamos que (1) para una sucesión de entrada a, la concatenación de λ y α, $\lambda\alpha$, y la concatenación de α y λ, $\alpha\lambda$, ambas levaran a la máquina de estado finito al mismo estado como lo hace α, y (2) si el estado inicial de una máquina de estado finito es un estado de aceptación, entonces λ está entre las sucesiones aceptadas por la máquina de estado finito.

El modelo de una máquina de estado finito puede extenderse, tal que permita la inclusión de flechas λ, como se muestra en la figura 7P. 14*a*. Notemos que la máquina de estado finito puede ir desde el estado *B* hasta el estado C "gratis".

a) Para la máquina de estado finito de la figura 7P. 14*b*, ¿cuáles de las siguientes sucesiones son aceptadas por la máquina de estado finito: 101, 10101, 110, 11010, 11001? Describa con palabras el conjunto de sucesiones aceptadas por la máquina.

b) La máquina de estado finito de la figura 7P. 14*b* se obtiene a partir de la máquina de la figura 7P. 14*a* por fusión de los estados *B* y *C*. Demuestre que las dos máquinas no aceptan el mismo conjunto de sucesiones a partir de un ejemplo de sucesiones que son aceptadas por una pero no por la otra.

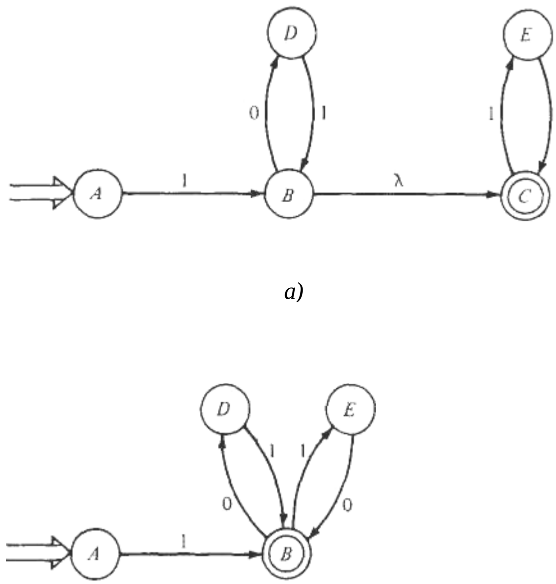


Figura 7P.14 *b}*

c) ¿Cómo puede transformarse la máquina de la figura 7P. 14*a* en una máquina que acepte el mismo conjunto de sucesiones sin usar flechas λ? Establezca un procedimiento general para tal transformación.

7.26 Sean L_1 y L_2 lenguajes de estado finito. Demuestre quees $L_1 \cup L_2$ también un lenguaje finito.

7.27 Sea *L* un lenguaje. Si L^R está dado por

$$L^R = \{x^R \mid x \in L\}$$

donde x^R denota la inversa de la sucesión *x*. Demuestre que si *L* es un lenguaje de estado finito, también lo es L^R.

7.28 Sea L un lenguaje que está especificado por una gramática en la cual las producciones son de la forma $A \to a$ y $A \to Ba$, donde A y B son no terminales y a es un terminal. Demuestre que L es un lenguaje de estado finito.

Sugerencia: utilice el resultado del problema 7.27.

7.29 Sea L un lenguaje especificado por una gramática en la cual las producciones son de la forma $A \to \gamma$ y $A \to \gamma B$ donde A y B son no terminales y y es una cadena de terminales. Demuestre que L es un lenguaje de estado finito.

CAPÍTULO

OCHO

Análisis de algoritmos

8.1 INTRODUCCIÓN

Un *algoritmo* es una especificación paso por paso sobre cómo realizar una cierta tarea. Aunque lo siguiente suena un poco técnico, una receta para preparar una comida es un algoritmo; también la hoja de instrucciones que acompaña a una bicicleta recién comprada, contiene un algoritmo para ensamblarla. Tenemos interés en estudiar el diseño y análisis de algoritmos por una razón. Al usar una computadora para realizar cualquier tarea, debemos especificar los pasos que deberá llevar a cabo, esto es, el algoritmo para realizar la tarea. De hecho, el lector ya ha visto varios algoritmos. En la sección 5.5, presentamos un algoritmo para determinar el paseo más corto entre dos vértices en un grafo pesado; en la sección 5.8, presentamos un algoritmo para determinar el itinerario de un agente viajero en un grafo pesado; en la sección 6.7 presentamos dos algoritmos diferentes para determinar un árbol generador mínimo de un grafo, y en la sección 8.8, también presentamos un algoritmo para determinar un flujo máximo en una red de transporte.

Ya sea que hayamos diseñado un algoritmo nosotros mismos o que nos presenten un algoritmo diseñado por otra persona, existen varias consideraciones importantes acerca del algoritmo. Primero, el algoritmo debe ser correcto -debe realizar su tarea asignada en forma correcta. Por ejemplo, un algoritmo para determinar un paseo más corto entre dos vértices dados debe producir efectivamente un paseo entre estos dos vértices y además, el paseo más corto entre los dos vértices. Segundo, nos gustaría conocer qué tan correcto es el resultado producido por el algoritmo. Esta pregunta es muy significativa aunque el algoritmo no prometa producir el mejor resultado posible. En caso de encontrar el paseo más corto entre dos vértices, y una vez que se ha comprobado que el algoritmo produce el paseo más corto entre los dos vértices, desde luego que el resultado es uno de los mejores posibles. Por otro lado, consideremos el algoritmo del vecino más cercano para el problema del agente viajero presentado en la sección 5.8. Puesto que el algoritmo no siempre produce el mejor resultado posible, lo deseable es tener la capacidad de evaluar la validez de su resultado. Recordemos que la longitud del itinerario producido por el algoritmo del vecino más cercano puede ser medida con respecto al itinerario más corto posible, como se mostró en el teorema 5.6. Como otro ejemplo, recordemos el algoritmo de programación de tareas presentado en la sección 4.7. Ahí, de nuevo, el algoritmo no siempre produce la mejor programación posible. No obstante, nos fue garantizado que el tiempo de ejecución total de acuerdo con la programación producida por nuestro algoritmo, nunca excederá el doble del menor tiempo de ejecución posible (teorema 4.2). Tercero, queremos determinar "el costo" de ejecución del algoritmo. Puesto que un algoritmo en verdad produce el resultado deseado, queremos conocer el costo de obtener el resultado. La medida del costo de ejecución de un algoritmo que comúnmente se usa más es la cantidad de tiempo que se lleva. No obstante, también existen otras medidas como el espacio de memoria requerido para ejecutar el algoritmo.

El estudio de los diferentes aspectos del diseño y análisis de algoritmos de cómputo es un tópico que debemos estudiar en este momento. Por otro lado, ya hemos dado al lector un bosquejo del material. En particular, como se señaló, el lector ya se ha visto ante algunas consideraciones sobre la exactitud y funcionalidad de los algoritmos en diversas ocasiones. En este capítulo, presentaremos algunos conceptos ligados con el costo de ejecutar un algoritmo, no sólo como una introducción al campo del análisis de algoritmos sino también como evidencia de que las matemáticas que hemos aprendido, efectivamente nos ayudarán a encarar muchos problemas.

8.2 COMPLEJIDAD TEMPORAL DE LOS ALGORITMOS

Como se dijo, nos gustaría determinar el costo de la ejecución de un algoritmo dado. Es obvio, el tiempo que toma ejecutar un algoritmo es una de las medidas más importantes del costo de ejecución. Para el resto de este capítulo, nos restringiremos a la medición del tiempo que toma la ejecución de un algoritmo, el cual también es conocido como la *complejidad temporal* del algoritmo.

Iniciemos con un ejemplo simple. Tenemos n números que son almacenados en n registros $x_1, x_2, \ldots, x_n$. Un número almacenado en un registro será referido como el contenido del registro. Sin pérdida de generalidad, supongamos que estos números son distintos. Queremos diseñar un algoritmo para determinar el mayor de estos n números. Podemos usar el siguiente algoritmo, al cual nos referiremos como el algoritmo MAYOR1.

Algoritmo MAYOR1

1. En principio, coloque el número del registro x_1 en un registro llamado *máx*.
2. Para $i = 2, 3, \ldots, n$, haga lo siguiente: compare el número en el registro x_i con el número en el registro *máx*. Si el número en x_i es mayor que el número en *máx*, cambie el número en el registro x_i al registro *máx*. En otro caso no haga nada.

3. Por último, el número en el registro *máx* es el mayor de los *n* números que hay en los registros $x_1, x_2, \ldots, x_n$.

No es difícil para el lector convencerse de que el algoritmo coloca efectivamente el mayor de los *n* números que hay en los registros $x_1, x_2, \ldots, x_n$ en el registro *máx*. Daremos, no obstante, una demostración formal de la efectividad del algoritmo.

DEMOSTRACIÓN Queremos demostrar mediante inducción la siguiente proposición: para cualquier i, $2 \leqq i \leqq n$, después de la ejecución del paso 2 el número en *máx* es el mayor de entre los números que están en los registros $x_1, x_2, \ldots, x_i$.

1. *Base.* Para $i = 2$, el número en *máx* es el mayor que x_1 y x_2.
2. *Paso de inducción.* Supongamos que la proposición es verdadera para $i = k$; esto es, el número en *máx* después de la ejecución del paso 2 para $i = k$ (pero antes de la ejecución del paso 2 para $i = k + 1$) es el mayor de entre los números en los registros $x_1, x_2, \ldots, x_k$. La ejecución del paso 2 para $i = k + 1$ compara el número en el registro x_{k+1} con el mayor de entre los números en los registros $x_1, x_2, \ldots, x_k$. Por tanto, el mayor de estos dos números será el mayor de entre los números en los registros $x_1, x_2, \ldots, x_k, x_{k+1}$.

Así, podemos concluir que el algoritmo MAYOR1 en verdad coloca en el registro *máx* al mayor de los *n* números almacenados en los registros $x_1, x_2, \ldots, x_n$. □

En cuanto al tiempo que toma ejecutar el algoritmo, observemos que cada uno de los números en los registros $x_2, x_3, \ldots, x_n$ es comparado una vez con el número del registro *máx*. Si denotamos con *c* el tiempo que toma comparar dos números, el tiempo total para las comparaciones es $(n - 1)c$. Desde luego, si somos extremadamente meticulosos, debemos notar que según sea el resultado de cada comparación, podemos necesitar o no, colocar un nuevo número en el registro *máx*, lo cual puede tomar un tiempo adicional. Sin embargo, tomar en cuenta esta variación nos conduciría a un análisis mucho más complicado, puesto que el número de veces que necesitamos para colocar un número nuevo en el registro *máx*, depende de la distribución de los *n* números en los registros $x_1, x_2, \ldots, x_n$. Por tanto, suponemos que el tiempo aplicado en colocar un número en el registro *máx* es insignificante o ha sido incluido en la cantidad c. En consecuencia, decimos que el tiempo total que toma ejecutar el algoritmo es $(n - 1)c$. Debido a que *c* es una cantidad que depende de la computadora que se use, con frecuencia decimos que el tiempo que toma ejecutar el algoritmo es proporcional a n - 1.

Ahora presentamos otro algoritmo para determinar el mayor de *n* números en los registros $x_1, x_2, \ldots, x_n$ al cual nos referiremos como el algoritmo MAYOR2.

Algoritmo MAYOR2

1. Haga lo siguiente para $i = 1, 2, \ldots, n - 1$: compare los números en los registros x_i y x_{i+1}. Coloque el mayor de los dos números en el registro x_{i+1} y el menor en el registro x_i.
2. Por último, el número en el registro x_n es el mayor de los *n* números.

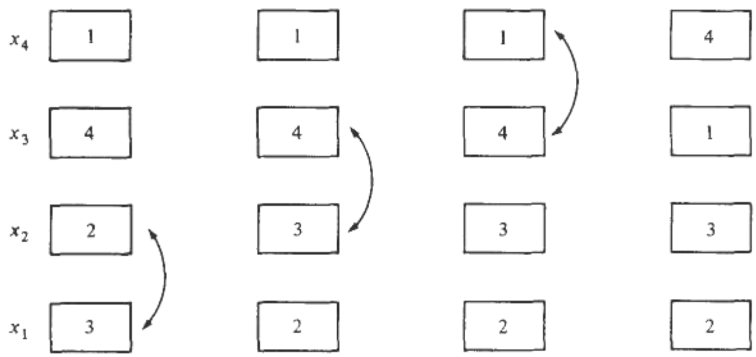


Figura 8.1

Primero presentamos un ejemplo ilustrativo que muestra cómo trabaja el algoritmo. La figura 8.1 muestra los pasos.

Para confirmar que el algoritmo es efectivamente correcto, exhortamos al lector a realizar una demostración por inducción de que para cualquier i, $1 \leqq i \leqq n - 1$, después de la ejecución del paso 1, el número en el registro x_{i+1} es el mayor de entre los i + 1 números en los registros $x_1, x_2, \ldots, x_i, x_{i+1}$. En cuanto a la complejidad temporal del algoritmo, que éste también es proporcional a n - 1 ya que realiza $n - 1$ comparaciones de números (el paso 1 es ejecutado n - 1 veces, y se hace una comparación cada vez).

En este momento deseamos señalar una ligera diferencia entre los dos algoritmos. El algoritmo **MAYOR1** no perturba el contenido de los n registros y siempre coloca al número mayor en el registro *máx*. Por otro lado, el algoritmo **MAYOR2** no sólo coloca al número mayor en el registro x_n sino que además rearregla el contenido de los otros registros como lo muestra el ejemplo anterior. Una razón para introducir un segundo algoritmo es mostrar al lector que algoritmos diferentes pueden solucionar el mismo problema. En el presente caso, sucede que los dos algoritmos tienen la misma complejidad temporal. Como veremos después, algoritmos diferentes que resuelven el mismo problema pueden tener complejidades temporales bastante diferentes. De la misma manera, algoritmos distintos pueden tener diferentes "efectos secundarios" aun cuando produzcan el mismo resultado. En este caso, el algoritmo **MAYOR1** no rearregla los números en los registros $x_1, x_2, \ldots, x_n$, en tanto que el algoritmo **MAYOR2** sí lo hace.

Ahora consideremos otro ejemplo. Por el ordenamiento de los n números almacenados en los registros $x_1, x_2, \ldots, x_n$, entendemos un rearreglo tal que los contenidos rearreglados en los registros $x_1, x_2, \ldots, x_n$, se encuentren en orden ascendente. Un algoritmo de ordenamiento, conocido como *ordenamiento burbuja*, funciona de la siguiente manera.

Algoritmo ORDENAMIENTO BURBUJA

1. Haga lo siguiente para $i = n, n-1, n-2, \ldots, 4, 3, 2$: utilice el algoritmo **MAYOR2** para colocar en el registro x_i al mayor de los i números en los registros $x_1, x_2, \ldots, x_i$.
2. Por último, los números en los registros $x_1, x_2, \ldots, x_n$ se encuentran en orden ascendente.

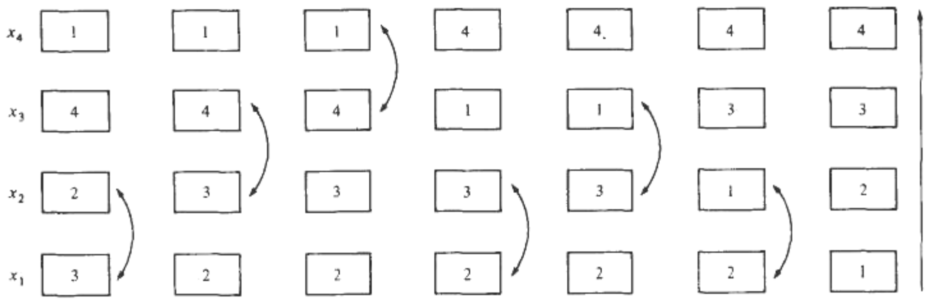


Figura 8.2

Para aseverar que el algoritmo ORDENAMIENTO BURBUJA en verdad ordena los *n* números en los registros $x_1, x_2, \ldots, x_n$, notamos que para *i* = *n,* la ejecución del paso 1 coloca el mayor de los *n* números en el registro x_n; para *i* = *n* - 1, la ejecución del paso 1 coloca el segundo número más grande de los *n* números en el registro x_{n-1}, etcétera. Por último, para *i* = 2, la ejecución del paso 1 coloca el (*n* - l)-ésimo más grande (esto es, el *segundo* más pequeño) de los *n* números en el registro x_2. Por tanto, el número que queda en el registro x_1 es el menor de los *n* números. Así, podemos concluir que el algoritmo sí rearregla los contenidos de los registros $x_1, x_2, \ldots, x_n$ en orden ascendente. La figura 8.2 muestra los pasos de un ejemplo.

Consideremos la complejidad temporal del algoritmo. La primera ejecución del paso 1 utiliza M - 1 comparaciones, la segunda ejecución del paso 1 utiliza *n* - 2 comparaciones, . . . , y la *(n* - l)-ésima ejecución del paso 1 utiliza una comparación. Así, el número total de comparaciones utilizadas es

$$(n-1)+(n-2)+\cdots+2+1=\frac{n(n-1)}{2}$$

En consecuencia, concluimos que la complejidad temporal del algoritmo es proporcional a $n(n-1)/2$.

8.3 ALGORITMO DEL PASEO MAS CORTO

Presentamos otro ejemplo que ilustra cómo analizar un algoritmo para determinar su complejidad temporal. Examinemos con minuciosidad el algoritmo que encuentra el paseo más corto de un vértice a *a* un vértice *z* en un grafo pesado presentado en la sección 5.5. Repetimos el algoritmo a continuación.

Algoritmo MASCORTO

1. Primero, sean $P = \{a\}$ y $T = V - \{a\}$. Para todo vértice *t* en *T*, sea $l(t) = w(a, t)$.
2. Seleccione el vértice de *T* que tiene el menor índice con respecto a *P*. Denote como *x* ese vértice.

3. Si x es el vértice que deseamos alcanzar desde a, pare. Si no lo es, sean $P' = P \cup \{x\}$ y $T' = T - \{x\}$. Para todo vértice t en T', calcule su índice con respecto a P' de acuerdo con (5.1).
4. Repita los pasos 2 y 3, use a P como P y a T' como T

El algoritmo comienza con $P = \{a\}$, incrementa el conjunto P un vértice a la vez, y termina cuando el conjunto P contiene al vértice z. Denotemos por n al número de vértices en el grafo. Así, el número máximo de iteraciones (de veces que se repiten los pasos 2 y 3) que realizará el algoritmo es n - 1. Durante cada iteración, necesitamos seleccionar el vértice de T que tiene el menor índice y recalcular los índices de los vértices en T'. Más específico, durante la primera iteración, necesitamos seleccionar un vértice de los n -1 vértices y recalcular los índices de los n - 2 vértices. Durante la segunda iteración, necesitamos seleccionar un vértice de los n - 2 vértices y recalcular los índices *den-3* vértices, y así sucesivamente. No obstante, siendo generosos, digamos que durante cada iteración seleccionamos un vértice de entre (no más de) n - 1 vértices y recalculamos los índices de (no más de) n - 1 vértices. Observemos que mediante un algoritmo similar al algoritmo **MAYOR1**, podemos seleccionar al menor de n - 1 números en un tiempo proporcional a n - 1; además, para calcular el índice de un vértice de acuerdo con (5.1) se requiere un tiempo constante que es independiente del número de vértices del grafo. Así, la complejidad temporal para recalcular los índices de n - 1 vértices también es proporcional a n - 1. En consecuencia, la complejidad temporal de los cálculos realizados en cada iteración es proporcional a $(n-1)$.†

Como que habrá a lo más n -1 iteraciones, concluimos que la complejidad temporal del algoritmo **MÁSCORTO** es proporcional a $(n-1)^2$. Cuando n es grande, la diferencia entre n y n - 1 se vuelve insignificante (¿realmente le importaría poseer mil millones de dólares o mil millones y un dólares?, ¿le importaría que un programa de computadora tarde 10 000 segundos o 10 001 segundos en ejecutar un programa?). De hecho, cuando n es grande, la diferencia entre n^2 y $(n-1)^2$ también se vuelve insignificante. Así, a menudo decimos que la complejidad temporal del algoritmo **MÁSCORTO** es proporcional a n^2 (analizaremos este resultado con más detalle en la sección 9.3).

8.4 COMPLEJIDAD DE LOS PROBLEMAS

Regresemos al problema simple de encontrar el mayor de n números almacenados en los registros $x_1, x_2, \ldots, x_n$. En la sección 8.2, vimos dos algoritmos que tienen una complejidad temporal proporcional a $n-1$. Algo que nos podríamos cuestionar de inmediato es: ¿podemos encontrar otro algoritmo con una complejidad temporal menor? Existen dos posibilidades. Supongamos que después de un esfuerzo considerable no logramos obtener

† Por supuesto, el lector debe comprender que si la complejidad temporal es proporcional a n - 1, también es proporcional a 2(n - 1); esto último es ligeramente redundante.

un algoritmo de complejidad temporal menor. ¿Qué podemos concluir?, ¿somos incompetentes o no existe un algoritmo con complejidad temporal menor? Por otro lado, si descubrimos un algoritmo con complejidad temporal menor; podríamos preguntarnos: ¿es posible hacerlo aún mejor? Entonces, sería deseable determinar la *complejidad temporal de un problema,* la cual se define como la complejidad temporal del "mejor" algoritmo posible para resolver el problema. En otras palabras, debido a la complejidad temporal del problema, debemos saber que existe un algoritmo que soluciona el problema de tal complejidad temporal, y que no existe un algoritmo con menor complejidad temporal que también solucione el problema. En la práctica, determinar la complejidad temporal de un problema es una tarea bastante difícil. En muchos casos, tan sólo nos es posible determinar una cota superior y una cota inferior para la complejidad temporal del problema. Una cota superior significa que un mejor algoritmo posible para solucionar el problema tendrá una complejidad temporal menor o igual a la cota. Una manera común para establecer una cota superior es simplemente diseñar un algoritmo que solucione el problema. Es claro que la complejidad temporal de cualquier algoritmo que solucione el problema es una cota superior para la complejidad temporal del problema. Por otro lado, para determinar una cota inferior debemos establecer que ningún algoritmo para solucionar el problema tendrá una complejidad menor que ésta.† En las contadas ocasiones en que la cota superior es igual a la cota inferior, hemos determinado la complejidad del problema. Además, si en verdad hemos diseñado un algoritmo para establecer una cota superior, este algoritmo es el mejor algoritmo posible. Ahora presentaremos dos ejemplos.

Ante el problema de determinar al mayor de n números dados, hemos establecido una cota superior para la complejidad del problema como n - 1 pasos de comparación, al demostrar un algoritmo para solucionar el problema que utiliza n - 1 pasos de comparación. Ahora deseamos establecer que n - 1 pasos de comparación también es una cota inferior para la complejidad del problema. En otras palabras, queremos mostrar que es imposible diseñar un algoritmo para solucionar el problema que use menos de n - 1 pasos de comparación. Supongamos un algoritmo que pretende solucionar nuestro problema usando menos de n - 1 pasos de comparación. Construyamos un grafo con n vértices $v_1, v_2, \ldots, v_n$ que corresponden a los n números. Si el algoritmo compara dos números que corresponden a los dos vértices v_i y v_j, habrá una arista que una a los vértices v_i y v_j. En consecuencia, si el algoritmo usa menos de n - 1 pasos de comparación, el grafo tendrá menos de n - 1 aristas. De acuerdo con nuestro análisis de la sección 6.1, el grafo debe ser un grafo no conexo. Esto significa que existen dos subconjuntos disjuntos de números entre los n números, de manera que ningún número en un subconjunto es comparado con algún número del otro subconjunto. Si éste es el caso, no importa cuantos números en cada subconjunto sean comparados uno con otro, el algoritmo no puede determinar cuál de los dos números mayores de los subconjuntos es el mayor. Así, el algoritmo no puede ser el algoritmo correcto que determine al mayor de n números.

Acerca de la complejidad del problema de determinar al menor de n números, hagámonos una pregunta trivial: ¿es obvio, mediante un argumento similar al anterior, que podemos concluir que la complejidad del problema también es proporcional a $n - 1$?

---

† Una vez más, ¿cómo demostramos que ningún algoritmo será mejor?

Ahora preguntémonos acerca de la complejidad del problema de determinar tanto el mayor como el menor de n números dados. En principio, estableceremos una cota superior para la complejidad del problema mediante el diseño de un algoritmo que solucione el problema. Consideremos a los n números almacenados en los registros $x_1, x_2, \ldots, x_n$. Podemos usar ya sea el algoritmo **MAYOR1** o bien el algoritmo **MAYOR2** para determinar al mayor de los n números. Después, usamos un algoritmo "similar" para determinar al menor de los $n - 1$ números restantes. Es claro que este algoritmo posee una complejidad temporal proporcional a

$$(n-1) + (n-2) = 2n - 3$$

Portante, 2w - 3 es una cota superior para la complejidad del problema. Resulta que existe un algoritmo más eficiente para este problema. En otras palabras, existe una cota superior más estricta para la complejidad del problema. Por simplicidad en la presentación, supongamos que n es un número par.† Sean $x_1, x_2, \ldots, x_n$ los n registros en los cuales están almacenados los n números. Consideremos el siguiente algoritmo.

Algoritmo MAYORMENOR

1. Para $i = 1, 2, \ldots, n/2$, compare los dos números de los registros x_i y $x_{i+(n/2)}$, y coloque al menor en el registro x_i y al mayor en el registro $x_{i+(n/2)}$.
2. Use el algoritmo **MAYOR1** para determinar al mayor de los $n/2$ números en los registros $x_{(n/2)+1}, x_{(n/2)+2}, \ldots, x_n$. Éste es el mayor de los n números dados.
3. Use un algoritmo similar a **MAYOR1** para determinar al menor de los $n/2$ números en los registros $x_1, x_2, \ldots, x_{n/2}$. Éste es el menor de los n números dados .

En principio necesitamos convencer al lector de que el algoritmo sí determina el mayor y el menor de los n números dados. Después del paso 1 el mayor de los n números está en uno de los registros $x_{n/2}, x_{(n/2)+1}, \ldots, x_n$, y el menor en uno de los registros $x_1, x_2, \ldots, x_{n/2}$ [un número en el registro x_i, $1 \leqq i \leqq n/2$, no puede ser el mayor de los n números debido a que es menor (al menos) que uno de los n números. De modo similar, un número en el registro x_i, $n/2+1 \leqq i \leqq n$, no puede ser el menor de los n números debido a que es mayor (al menos) que uno de los n números]. Por tanto, el paso 2 sí determina al mayor de los n números y el paso 3 determina al menor.

En cuanto a la complejidad temporal del algoritmo, el paso 1 requiere $n/2$ pasos de comparación; el paso 2 requiere $(n/2)$ -1 pasos de comparación, y el paso 3 requiere $(n/2)$ - 1 pasos de comparación. Así, el número total de pasos de comparación es $(3n/2) - 2$, lo cual es una mejora sobre el algoritmo sugerido al inicio.

---

† Dejamos al lector el caso donde n es impar.

Concluiremos nuestro análisis (y dejaremos al lector el reto de intentar mejorar aún más el resultado del párrafo anterior) mostrando que $(3n/2) - 2$ es una cota inferior para la complejidad del problema. Emplearemos lo que se conoce como un *argumento del adversario* el cual cuenta los pasos de un algoritmo con resultados desfavorables. Mostraremos que sin importar el diseño del algoritmo, el argumento del adversario forzará al algoritmo a usar al menos $(3n/2) - 2$ pasos de comparación para determinar tanto al mayor como al menor de los n números dados. Supongamos que nos presentan un algoritmo que soluciona nuestro problema. Sean $a_1, a_2, \ldots, a_n$ los n números dados. En un momento cualquiera durante la ejecución del algoritmo, denotamos como A al conjunto de números que aún no han sido comparados con cualquier otro número; denotamos como B al conjunto de números que se ha confirmado que no son el mayor, y denotamos como C al conjunto de números que se ha confirmado que no son el menor. Al inicio de la ejecución del algoritmo, $A = \{a_1, a_2, \ldots, a_n\}$, $B = \phi$ y $C = \phi$. Al final de la ejecución, $|A| = 0$, $|B| = n - 1$ y $|C| = n - 1$. Observemos que un número podría aparecer tanto en B como en C. Por otro lado, es claro que *Ay B* son disjuntos, como lo son *A y Cha* idea del argumento del adversario es que siempre que el algoritmo compare dos números, el argumento del adversario informará el resultado del algoritmo. El argumento del adversario selecciona los resultados de tal manera que el algoritmo no podrá determinar al mayor y al menor de los n números hasta que haya efectuado al menos $(3n/2) - 2$ pasos de comparación. En otras palabras, el algoritmo no podrá construir los conjuntos B y C tal que $|B| = n - 1$ y $|C| = n - 1$, hasta que haya realizado al menos $(3n/2) - 2$ pasos de comparación. Desde luego, la respuesta del argumento del adversario siempre deberá ser consistente (esto es, si el argumento del adversario indica al algoritmo que $a_i < a_j$ y $a_j < a_k$, entonces debe responder con $a_i < a_k$ siempre que el algoritmo pregunte por el resultado de comparar a_i y a_k. El argumento del adversario utiliza las siguientes reglas:

1. Si el algoritmo compara dos números a_i y a_j tales que a, esté en B pero no en C, entonces $a_i < a_j$ (si tanto a_i como a_j- están en B pero no en C, el resultado es arbitrario en tanto sea consistente).
2. Si el algoritmo compara dos números a_i y a_j tales que a_i esté en C pero no en B, entonces $a_i > a_j$ (si tanto a_i como a_j están en C pero no en B, el resultado es arbitrario en tanto sea consistente).
3. Para cualquier otra comparación, el resultado es arbitrario siempre que sea consistente.

Primero observemos que la respuesta del argumento del adversario siempre es consistente. Si el número a_i está en B pero no en C, esto significa que a_i nunca ha sido comparado con un número menor que a_i. Por tanto, la regla 1 nunca dará una respuesta inconsistente. Si el número a_i está en C pero no en B, esto significa que a, nunca ha sido comparado con un número mayor que a_i. Así, la regla 2 nunca dará una respuesta inconsistente.

Segundo, notemos que al comparar dos números en A, ambos deben eliminarse de A, el menor de los dos debe añadirse a B, y el mayor añadirse a C. Así, tanto | B | como | C | serán incrementados por 1. Cualquier otra comparación causará *bien que \ B* | se incremente por 1 *o bien que \ C* | se incremente por 1. En consecuencia, para incrementar | B | desde 0 hasta

n - 1 y para incrementar | C | desde 0 hasta n- 1, lo mejor que cualquier algoritmo puede hacer es usar $n/2$ comparaciones para los n números que estaban inicialmente en A y usar otras $2(n-1)-n$ comparaciones, de manera que los tamaños de | B | y | C | lleguen a n - 1 al final. Así, el número total de comparaciones es al menos

$$\frac{n}{2}+[2(n-1)-n]=\frac{3n}{2}-2$$

Podemos concluir que (3n/2) - 2 es una cota inferior para la complejidad temporal del problema de encontrar al mayor y al menor de entre n números.

8.5 PROBLEMAS TRATABLES Y NO TRATABLES

Una vez que hemos comprendido la noción de la complejidad temporal de los problemas, podemos adquirir un sentido intuitivo sobre la noción de la dificultad o facilidad computacional de los problemas. Cuando empezamos a usar las computadoras digitales, de inmediato aprendimos que una computadora puede llevar a cabo operaciones aritméticas y lógicas a una velocidad muy rápida, a saber, millones o más, operaciones por segundo. Debido a que casi todos los problemas que encontramos en computación digital son finitos por naturaleza, parece ser que no hay razón para no poder solucionar cualquier problema. Entonces, no existe problema que no pueda ser resuelto en su totalidad. Regresemos al problema del agente viajero analizado en la sección 5.8. Si en total hay (n - 1)! posibles itinerarios (¿por qué?), el siguiente algoritmo seguramente solucionará el problema: calcule el costo de cada uno de los (n - 1)! itinerarios y escoja el que tiene menor costo.

Así, si nuestro problema tiene cinco ciudades, simplemente podemos examinar todos los $4! = 24$ itinerarios; si tiene 10 ciudades, podemos examinar todos los 9! = 362 880 itinerarios, y si tiene 70 ciudades, podemos examinar todos los 69! itinerarios. La única falla en esta solución es que n! se vuelve muy grande inclusive para n moderadamente grande. Tomemos como ejemplo el problema con 70 ciudades, y supongamos que nuestra computadora puede examinar 10^{10} itinerarios por segundo. Dado que

$$69! = 1.71122 \times 10^{98}$$

le tomará a la computadora 1.71122×10^{88} segundos, cerca de 5.42626×10^{78} siglos para examinar todos los itinerarios, ¡y hasta entonces seleccionaremos el de menor costo!

Examinemos otro problema. Supongamos n enteros $a_1, a_2, \ldots, a_n$ y una constante K. Queremos encontrar un subconjunto de enteros $a_{i1}, a_{i2}, \ldots$ tales que su suma sea menor o igual y tan cercana como sea posible a K. Por ejemplo, dados los 11 enteros

4,7,8,10,12,17,25,29,31,36

y K = 63, exhortamos al lector a seleccionar algunos de estos enteros de manera que su suma sea menor o igual a 63 y tan cercana como sea posible a 63. Este problema, conocido como

Tabla 8.1

	2	5	10	50	60	100
n	2	5	10	50	60	10^2
n^2	4	25	10^2	2.5×10^3	3.6×10^3	10^4
n^3	8	125	10^3	1.25×10^5	2.16×10^5	10^6
n^5	32	3125	10^5	3.12×10^8	7.78×10^8	10^{10}
2^n	4	32	10^3	1.13×10^{15}	1.15×10^{18}	1.27×10^{30}
3^n	9	243	5.9×10^4	7.18×10^{23}	4.24×10^{28}	5.15×10^{47}
$n!$	2	120	3.63×10^6	3.04×10^{64}	8.32×10^{81}	9.33×10^{177}

el *problema de la mochila*, proviene de la siguiente situación: tenemos una mochila de excursionista de tamaño K y n objetos de tamaño $a_1, a_2, \ldots, a_n$. Queremos guardar algunos objetos dentro de la mochila de manera que ocupen el mínimo espacio posible. Existe un algoritmo para solucionar el problema: para cada subconjunto de $\{a_1, a_2, \ldots, a_n\}$, calcule la diferencia entre la suma de los elementos en el subconjunto y K. Es obvio que el subconjunto que dé lugar a la menor diferencia no negativa es la solución. Dado que el conjunto $\{a_1, a_2, \ldots, a_n\}$ tiene 2^n subconjuntos, examinar todos los subconjuntos nos permite seleccionar al subconjunto que proporcione la menor diferencia no negativa. De esto se cumple que la complejidad temporal de tal algoritmo es proporcional a 2^n. Para $n = 100$, $2^n = 1.26765 \times 10^{30}$. Incluso si suponemos que nuestra computadora puede examinar 10^{10} subconjuntos por segundo, le tomará 4.01969×10^{12} años en completar el cálculo.

Observemos que n, n^2, n^3 y, para un k fijo, n^k, crecen lentamente conforme se incrementa n, en tanto que $n!$, 2^n y 3^n crecen con rapidez al incrementar n, como se muestra en la tabla 8.1 (en la sección 9.3 analizaremos con detalle la razón de crecimiento de funciones). Así, un algoritmo con una complejidad temporal que no crece más rápido que n^k es *eficiente*, en tanto que un algoritmo con una complejidad temporal que crece más rápido que n^k es *ineficiente*.† Así, los dos algoritmos presentados son ejemplos de algoritmos ineficientes. Por otro lado, los algoritmos presentados en la sección 8.2 para determinar al mayor de n números y el algoritmo del paseo más corto presentado en la sección 8.3 son ejemplos de algoritmos eficientes.

De esto se obtiene que un problema se considera *tratable* (computacionalmente sencillo) si puede ser resuelto mediante un algoritmo eficiente y *no tratable* (computacionalmente difícil) si no existe un algoritmo eficiente para solucionarlo. Para mostrar que un problema es tratable, sólo necesitamos demostrar un algoritmo eficiente para solucionar el problema. Así, ya hemos demostrado que los problemas de determinar al mayor de n números, ordenar n números y encontrar el paseo más corto entre dos vértices dados en un grafo, son problemas

† n es el tamaño de la muestra del problema que debe ser resuelto. Por ejemplo, en el caso de hallar al número más grande de entre un conjunto determinado de números, n es la cantidad de números que hay en el conjunto. Para el caso del ordenamiento, n es la cantidad de números que serán ordenados. En el caso de hallar el paseo más corto en un grafo, n es el número de vértices en el grafo.

tratables. Para señalar que un problema es no tratable, debemos mostrar que no existe un algoritmo eficiente para solucionarlo (en otras palabras, debemos mostrar una cota inferior para la complejidad del problema que crezca más rápido que n^k). En efecto, se ha demostrado que existen problemas que son no tratables, pero un análisis detallado de este tópico se encuentra fuera del alcance de este libro. Sin embargo, nuestro análisis nos conduce hacia uno de los más importantes problemas no-resueltos en la teoría de la ciencia de la computación. Existe una clase de problemas, que incluyen el problema del agente viajero y el problema de la mochila, para los cuales, hasta el momento no se conoce algún algoritmo eficiente. Por otro lado, no se ha demostrado que no puedan existir algoritmos eficientes para su solución. En otras palabras, no podemos confirmar si esta clase de problemas es tratable o no. Lo que hace a esta clase de problemas muy importante es que se ha demostrado que son todos tratables o todos no tratables. Esto es, si podemos encontrar un algoritmo eficiente para solucionar cualquier problema de esa clase, deberíamos entonces tener algoritmos eficientes para solucionar todos los problemas de esa clase. O si podemos demostrar que cualquier problema de esa clase es no tratable, demostramos que todos los problemas de esa clase son no tratables. Esta clase de problemas es conocida como la *clase de problemas NP completos [NP* es la abreviatura de *nondeterministic polynomial* (o polinomios no determinísticos). En la sección 8.6 encontrará algunas referencias].

8.6 NOTAS Y REFERENCIAS

Los libros de Knuth [8, 9, 10] presentan varios aspectos del análisis de algoritmos. Véase también Aho, Hopcroft y Ullman [1, 2], Baase [3], Horowitz y Sahni [6], Hu [7], Reingold, Nievergelt y Deo [11], y Sedgewick [12]. La noción de problemas tratable y no tratable es uno de los conceptos más importantes en la teoría de las ciencias de la computación. La noción de NP completos fue introducida por S. A. Cook [4]. La referencia más completa del tema es la obra de Garey y Johnson [5]. Consulte también las referencias generales citadas arriba.

1. Aho, A. V., .1. E. Hopcroft y .1. D. Ullman: *The Design and Analysis of Computer Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1974.
2. Aho, A. V., J. E. Hopcroñ y J. D. Ullman: *Data Structures and Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1983.
3. Baase, S.: *Computer Algorithms: Introduction to Design and Analysis,* Addison-Wesley Publi shing Company, Reading, Mass., 1977.
4. Cook, S. A.: "The Complexity of Theorem-proving Procedures", *Proceedings of the Third Annual ACMSymposium on the Theory of Computing,* 151-158, (1971).
5. Garey, M. R. y D. S. Johnson: *Computers and Intractability: A Guide to the Theory ofNP-Completeness,* Freeman, Nueva York, 1979.
6. Horowitz, E. y S. Sahni: *Fundamentals of Computer Algorithms,* Computer Science Press, Potomac, Md., 1978.
7. Hu, T. C: *Combinatorial Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1982.
8. Knuth, D. E.: *The Art of Computer Programming, Vol. I, Fundamental Algorithms,* 2d ed., Addison-Wesley Publishing Company, Reading, Mass., 1973.
9. Knuth, D. E.: *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1968.

10. Knuth, D. E.: *The Art of Computer Programming, Vol. 3, Sorting and Searching,* Addison-Wesley Publishing Company, Reading, Mass., 1973.
11. Reingold, E. M., J. Nievergelt y N. Deo: *Combinatorial Algorithms: Theory and Practice,* Prentice-Hall, Englewood Cliffs, N.J., 1977.
12. Sedgewick, R.: *Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1983.

PROBLEMAS

8.1 *a)* Diseñe un algoritmo que determine si n bolas de color situadas en n cajas son del mismo color. La operación básica es comparar dos bolas en cualquiera de dos cajas y encontrar si son del mismo color. Determine la complejidad del algoritmo en términos del número de operaciones básicas usadas. También determine la complejidad del problema.
b) Diseñe un algoritmo que determine si n bolas son de uno o de dos colores. ¿Cuál es la complejidad del algoritmo?

8.2 Dada una sucesión de n números 0 y 1, se desea organizados de manera que los números 0 queden agrupados a la izquierda y los números 1 a la derecha. La operación básica es comparar dos dígitos *adyacentes* e intercambiar sus posiciones, si se quiere. Diseñe un algoritmo y determine su complejidad.

8.3 Diseñe un algoritmo que seleccione el número más grande y el segundo más grande de n números. La operación básica es comparar dos números y determinar el más grande y el más pequeño de los dos. ¿Cuál es la complejidad del algoritmo?

8.4 Dado un árbol m-ario completo de altura h con un peso asignado a cada arista (por ejemplo, el árbol ternario de la figura 8P.1), diseñe un algoritmo que encuentre un paseo de peso mínimo de la raíz del árbol a cualquiera de sus hojas. ¿Cuál es la complejidad del algoritmo?

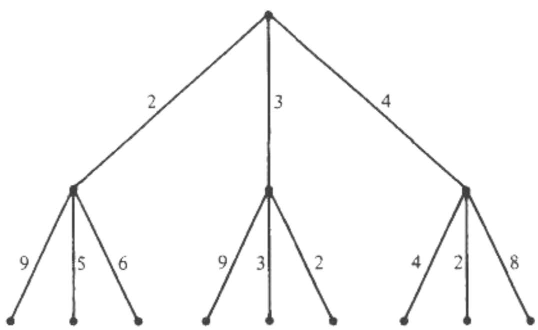


Figura 8P.1

8.5 ¿Cuál es la complejidad del algoritmo del vecino más cercano para el problema del agente viajero de la sección 5.8?

8.6 Examinemos el problema de ordenación de n números, $x_1, x_2, \ldots, x_n$. Sin pérdida de generalidad, podemos suponer que los n números son distintos. Considere una clase de algoritmos de ordenamiento en los que la operación básica sea comparar dos números y acomodarlos de acuerdo con el resultado de la comparación. Esto es, podemos comparar cualquier par de números x_i y x_j.
Si $x_i > x_j$, entonces un curso de acción será seguido. Sin embargo, si $x_i < x_j$, entonces otro curso de acción será seguido. Cualquier algoritmo puede ser convenientemente representado por un árbol binario, donde cada nodo interno corresponde a una operación de comparación y cada hoja corresponde a una salida final con el orden de los n números determinado por completo.

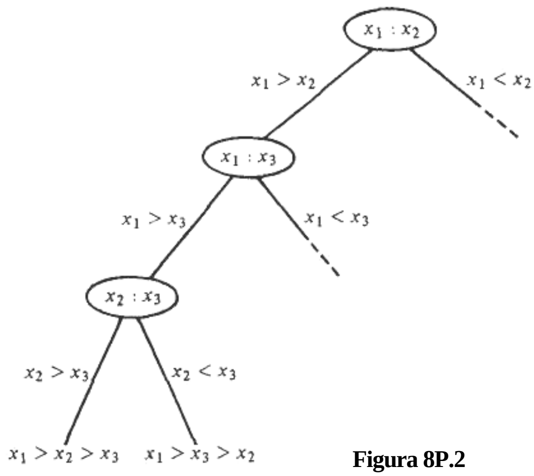


Figura 8P.2

a) La figura 8P.2 muestra parte de un árbol binario que describe un algoritmo para ordenar tres números. Complete la descripción del algoritmo. ¿Cuál es la complejidad del algoritmo, en términos del número de comparaciones hechas?

b) Diseñe un algoritmo para ordenar cuatro números y determine su complejidad.

c) Determine una cota inferior de complejidad de cualquier algoritmo que ordene *n* números en esta clase.

8.7 De entre *n* monedas, a lo más una de ellas es defectuosa (una moneda defectuosa puede ser más pesada o más ligera que una moneda en buen estado). También hay una fuente de monedas que se sabe que son buenas. Se nos da una balanza que nos permite comparar los pesos totales de las monedas colocadas en los dos lados de la balanza. Queremos diseñar algoritmos que determinen si es que todas las monedas están en buen estado, y si no, queremos identificar la moneda defectuosa y determinar si ésta es más pesada o más ligera que una en buen estado. En particular, investigaremos una clase de algoritmos que nos permitan tres maneras de proceder después de haber usado la balanza. Esto es, según sea el peso total de las monedas en el plato izquierdo de la balanza, mayor, igual o menor que el peso total de las monedas en el plato derecho, diferentes cursos de acción serán seguidos. La complejidad de tales algoritmos se medirá mediante el número de veces que la balanza sea usada.

a) Para $n = 1$, muestre un algoritmo que solucione el problema cuando la balanza se usa una vez, y concluya (trivialmente) que, en esta clase de algoritmos, la complejidad del problema en realidad es usar la balanza una vez.

b) Para $n = 4$, muestre un algoritmo que solucione el problema cuando la balanza se usa dos veces. Una manera conveniente para describir el algoritmo podría ser un árbol ternario en el cual cada uno de los nodos internos represente el uso por una vez de la balanza. Determine una cota inferior para la complejidad de todos los algoritmos en esta clase.

c) Demuestre que la cota inferior en la complejidad de todos los algoritmos en esta clase es usar la balanza *k* veces, donde *k* es el entero más pequeño tal que $n \leqq (3^k - 1)/2$.

d) ¿Puede encontrarse un algoritmo no bifurcable para $n = 4$ que también use la balanza sólo dos veces, es decir, un algoritmo en el cual las mismas comparaciones de pesos sean hechas para el segundo uso de la balanza sin tomar en cuenta el resultado del primer uso? ¿Puede encontrarse un algoritmo no bifurcable para $n = 13$ que use la balanza tres veces?, ¿para $n = (3^k - 1)/2$ que use la balanza *k* veces?

Sugerencia: esta parte es no trivial.

8.8 Dados los valores de $a_0, a_1, a_2, \ldots, a_n$ y x, queremos evaluar el polinomio

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

En forma más directa, podemos calcular el valor de cada término $a_i x^i$ y sumarlos. Ya que el producto $a_i x^i = a_i \cdot x \cdot x \ldots \cdot x$ puede calcularse en i operaciones de multiplicación, necesitamos un total de $n(n - \backslash)/2$ operaciones de multiplicación y n de suma. ¿Puede hacer algo mejor?

8.9 Considere el siguiente algoritmo que obtiene la suma de $x_1, x_2, \ldots, x_8$:

suma := x_1
de $i = 2$ hasta 8
suma := suma + x_i

La ejecución de este algoritmo requiere de siete unidades de tiempo, si suponemos que éste toma una unidad de tiempo para sumar dos números. Ahora, supongamos que tenemos un gran número de procesadores que pueden usarse para realizar la suma en paralelo. Podemos emplear el siguiente algoritmo, donde todas las operaciones de suma con la orden "hágase en paralelo", se realizarán en forma simultánea en distintos procesadores. En consecuencia, la ejecución de este algoritmo requiere tres unidades de tiempo.

1. Hágase en paralelo

$$x_1 := x_1 + x_2$$
$$x_3 := x_3 + x_4$$
$$x_5 := x_5 + x_6$$
$$x_7 := x_7 + x_8$$

2. Hágase en paralelo

$$x_1 := x_1 + x_3$$
$$x_5 := x_5 + x_7$$

3. suma \=$x_j + x_5$

Supongamos un modelo de cálculo en paralelo con un número infinito de procesadores que operan simultáneamente. En una unidad de tiempo, un procesador puede accesar un número fijo de registros, llevar a cabo una operación aritmética, lógica o de comparación, y almacenar los resultados en un número fijo de registros.

a) Mediante un modelo de cálculo en paralelo, diseñe un algoritmo similar al presentado para encontrar la suma de 2* números. ¿Cuál es la complejidad del algoritmo?

b) Diseñe el algoritmo en paralelo que determine el máximo de 2^k números. ¿Cuál es la com plejidad del algoritmo?

c) Diseñe un algoritmo en paralelo que determine el número de números positivos de entre 2* números. ¿Cuál es la complejidad del algoritmo?

8.10 *Unapila* es un arreglo de números en los nodos de un árbol binario tal que las siguientes condiciones se satisfacen:

Todas las hojas están en el más bajo o en el penúltimo nivel.
Todas las hojas del nivel más bajo están en las posiciones más cercanas a la izquierda del árbol.
El número en cualquier nodo interno es más grande que los números en sus dos hijos.

Por ejemplo, el árbol binario de la figura 8P.3a es una pila, mientras que el de la figura 8P.36 no lo es.

a) Dado un árbol binario con *n* números tales que las dos primeras condiciones de arriba se satisfacen, diseñe un algoritmo que convierta al árbol en una pila. ¿Cuál es la complejidad del algoritmo?

b) Suponga que nos han dado una pila con *n* números. Si reemplazamos el número de la raíz con el número de la hoja de más a la derecha en el nivel más abajo, ya no tendremos una pila (por ejemplo, la figura 8P.3c se obtiene a partir de la figura 8P.3a de esta manera). Diseñe un algoritmo para rearreglar un árbol en una pila. ¿Cuál es la complejidad del algoritmo?

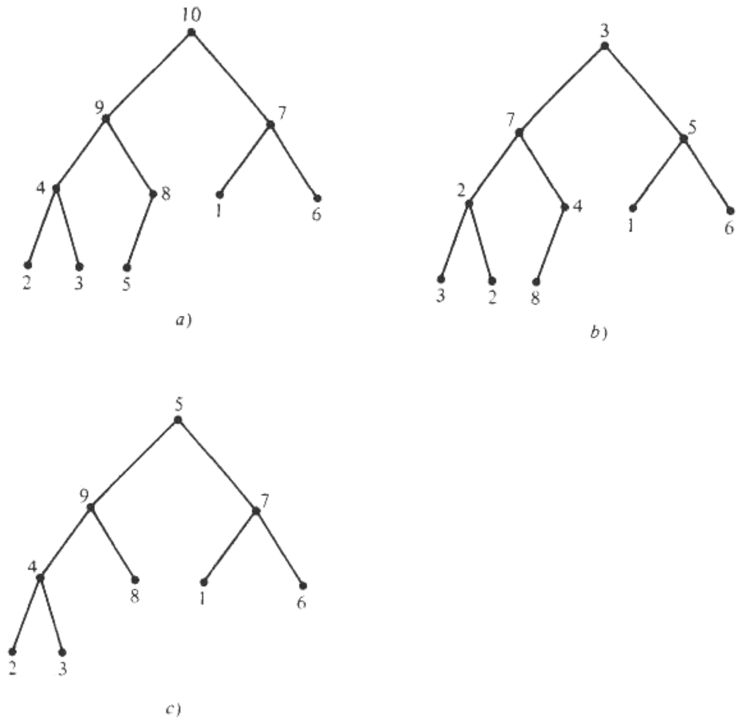


Figura 8P.3

c) Dada una pila de *n* números, ¿qué puede decir acerca del número de la raíz? Los algoritmos diseñados en los incisos *a)* y *b)* pueden combinarse para obtener un algoritmo de ordenamiento, ¿de qué manera? ¿Cuál es la complejidad del algoritmo de ordenamiento?

8.11 Estudiamos el problema de calcular x^n para x y n; supongamos que todos los resultados intermedios del cálculo están disponibles para nosotros.

a) Dado *x*, ¿se puede calcular x^{16} en cuatro multiplicaciones?

b) Dado *x*, ¿con cuántas multiplicaciones se puede calcular x^{15} y x^{23}?

c) Para un entero n, sea $b_k\, b_{k-1} \cdots b_1 b_0$ la representación binaria del número. Para cada b_i reemplazamos b_i por SX si $b_i = 1$, y reemplazamos b_i por S si $b_i = 0$. Por ejemplo, dado que la representación binaria para el número 29 es 11101, obtenemos la secuencia *SXSXSXSSX*. Al eliminar el primer *SX*, obtenemos la secuencia *SXSXSSX*. Si sustituimos cada *S* por un *cuadrado,* y cada *X* por *multiplicar por x,* obtenemos una secuencia de pasos *{cuadrado, multiplicar por x, cuadrado, multiplicar por x, cuadrado, cuadrado, multiplicar por x}*, donde *cuadrado* significa calcular el cuadrado del resultado actual, y *multiplicar por x* significa multiplicar el resultado actual por *x.* Demuestre que si se inicia con *x* como el resultado actual, para $n = 29$, esta secuencia de pasos efectivamente calculará x^{29}.

d) Muestre cómo se puede calcular x^{59} con el algoritmo presentado en el inciso *c)*.

e) Demuestre que el algoritmo presentado en el inciso *c)* es en efecto correcto.

f) Dado un entero *n,* una cadena de adición para *n* es una sucesión de enteros

$$a_0\ a_1\ a_2\ \cdots\ a_m$$

tal que $a_0 = 1$, $a_m = n$ y $a_i = a_j + a_k$ para $k \leqq j < 1$. Por ejemplo,

1 2 3 5 7 14 28 29

y

1 2 3 6 7 14 28 29

ambas son cadenas de adición para 2. ¿Cuál es la relación entre las cadenas de adición y la evaluación de las potencias de *x*?

g) Determine las cadenas de adición para 19, 33, 46, 79 y 87. ¿En cuántas multiplicaciones pueden evaluarse: x^{19}, x^{33}, x^{46}, x^{79}, x^{87}?

CAPÍTULO

NUEVE

Funciones numéricas discretas y funciones generatrices

9.1 INTRODUCCIÓN

Recordemos que una función es una relación binaria que asigna a cada elemento en el dominio un valor único, el cual es un elemento del rango. En éste y el siguiente capítulo, estudiaremos una clase de funciones cuyo dominio es el conjunto de los números naturales y cuyo rango es el conjunto de los números reales. Nos referiremos a estas funciones como las *funciones numéricas discretas o funciones numéricas.* Las funciones numéricas son de particular interés para nosotros debido a que las encontraremos con frecuencia en la computación digital.

Para denotar las funciones numéricas usaremos letras minúsculas en negritas. Para una función numérica **a** usamos $a_0, a_1, a_2, \ldots, a_r, \ldots$ para denotar los valores de la función en $0, 1, 2, \ldots, r, \ldots$ †

En principio, podemos especificar una función numérica al listar exhaustivamente sus valores $(a_0, a_1, a_2, \ldots)$, pero en la práctica necesitamos usar una representación que no sea infinita. Así,

$$a_r = 7r^3 + 1 \qquad r \geqq 0$$

$$b_r = \begin{cases} 2r & 0 \leqq r \leqq 11 \\ 3^r - 1 & r \geqq 11 \end{cases}$$

$$c_r = \begin{cases} -4 & r = 3, 5, 7 \\ 0 & \text{en otro caso} \end{cases}$$

$$d_r = \begin{cases} 2 + r & 0 \leqq r \leqq 5 \\ 2 - r & r > 5 \quad r \text{ es impar} \\ 2/r & r > 5 \quad r \text{ es par} \end{cases}$$

† Esto es una notación mas simple que la que se dio en la sección 4.8, donde usamos a(0), a(l), a(2),..., **a**(r), ...

son ejemplos de especificaciones concisas para funciones numéricas. Cuando exista una expresión simple para el valor de una función numérica para cada valor de r tal como las a_r anteriores, también usaremos la notación

$$\mathbf{a} = 7r^3 + 1$$

Consideremos los siguientes ejemplos:

Ejemplo 9.1

Suponga que depositamos \$100 en una cuenta de ahorros a una tasa de interés del 7 por ciento anual. Al final del primer año, la cantidad total en la cuenta es \$107; al final del segundo año, la cantidad total es \$114.49; al final del tercer año, la cantidad total es \$122.50, etcétera. La cantidad en la cuenta al final de cada año puede ser descrita mediante una función numérica a, la cual puede especificarse como (100, 107, 114.49, 122.50,...) o como

$$a_r = 100(1.07)^r \qquad r \geqq 0$$

o también

$$\mathbf{a} = 100(1.07)^r$$

□

Ejemplo 9-2

Sea a_r la altitud de una aeronave a miles de pies en el r-ésimo minuto. Supongamos que la aeronave despega después de estar 10 minutos sobre tierra, asciende a una velocidad uniforme hasta una altitud de crucero de 30 000 pies en 10 minutos, comienza a descender uniformemente después de 110 minutos de tiempo de vuelo, y aterriza 10 minutos después. Luego tenemos que

$$a_r = \begin{cases} 0 & 0 \leqq r \leqq 10 \\ 3(r-10) & 11 \leqq r \leqq 20 \\ 30 & 21 \leqq r \leqq 120 \\ 3(130-r) & 121 \leqq r \leqq 130 \\ 0 & r \geqq 131 \end{cases}$$

□

9-2 MANIPULACIÓN DE FUNCIONES NUMÉRICAS

La *suma* de dos funciones numéricas es una función numérica cuyo valor en r es igual a la suma de los valores de las dos funciones numéricas en r. El *producto* de dos funciones numéricas es una función numérica cuyo valor en r es igual al producto de los valores de las dos funciones numéricas en r. Por ejemplo, consideremos las funciones a y b donde

$$a_r = \begin{cases} 0 & 0 \leqq r \leqq 2 \\ 2^{-r} + 5 & r \geqq 3 \end{cases}$$

y

$$b_r = \begin{cases} 3 - 2^r & 0 \leqq r \leqq 1 \\ r + 2 & r \geqq 2 \end{cases}$$

Si c es la suma de las dos funciones a y b, la cual denotaremos por a + b, entonces

$$c_r = a_r + b_r = \begin{cases} 3 - 2^r & 0 \leqq r \leqq 1 \\ 4 & r = 2 \\ 2^{-r} + r + 7 & r \geqq 3 \end{cases}$$

Sea d el producto de las funciones a y b, que denotaremos por ab, entonces

$$d_r = a_r b_r = \begin{cases} 0 & 0 \leqq r \leqq 2 \\ r \cdot 2^{-r} + 2^{-r+1} + 5r + 10 & r \geqq 3 \end{cases}$$

Por ejemplo, si a denota el ingreso mensual del marido y b el ingreso mensual de su esposa, entonces a + b será su ingreso mensual conjunto. Si a denota el balance de una cuenta de ahorros en cada mes y b la tasa de interés mensual, la cual fluctúa mensualmente, entonces ab será el interés ganado en cada mes.

Sea a una función numérica y a un número real. Con oca denotamos la función numérica cuyo valor en r es igual a a veces a_r A la función numérica α a la llamamos *versión escalada* de a con *factor de escala* α. Por ejemplo, sea a una función numérica con valor en r de $(1.07)^r$, entonces 100a es una función numérica cuyo valor en r es $100(1.07)^r$. Es obvio, si a describe la cantidad total de una cuenta de ahorros a través de los años, a partir de un depósito inicial de $ 1, entonces 100a describe la cantidad total en la cuenta para un depósito inicial de $100.

Para denotar una función numérica usamos |a| cuyo valor en r es igual a a_r si a_r es no-negativo, y es igual a —a_r si a_r es negativo. Por ejemplo, sea

$$a_r = (-1)^r \left(\frac{2}{r^2}\right) \qquad r \geq 0$$

si b es |a|, entonces

$$b_r = \frac{2}{r^2} \qquad r \geq 0$$

Por ejemplo, a es una función numérica donde a_r es la diferencia entre el voltaje de salida de una fuente de poder eléctrica, después de r horas de operación, y un voltaje nominal de referencia. Así, un a_r positivo significa que el voltaje después de r horas de operación está por arriba del voltaje de referencia, en tanto que un a_r negativo significa que el voltaje después de r horas de operación está por debajo del voltaje de referencia. Así, |a| nos da la desviación del voltaje de salida respecto al voltaje de referencia.

Sea a una función numérica e i un entero positivo. Usamos $S^i a$ para denotar una función numérica tal que su valor en r es 0 para $r = 0, 1, 2, \ldots, i - 1$ y es a_{r-i} para $r \geqq i$. Por ejemplo,

$$a_r = \begin{cases} 1 & 0 \leqq r \leqq 10 \\ 2 & r \geqq 11 \end{cases}$$

Si b es S^5a, entonces

$$b_r = \begin{cases} 0 & 0 \leqq r \leqq 4 \\ 1 & 5 \leqq r \leqq 15 \\ 2 & r \geqq 16 \end{cases}$$

Como otro ejemplo, si a describe la altitud de una aeronave como se señaló en un ejemplo anterior, entonces $S^{17}\mathbf{a}$ describe la altitud cuando el despegue se retrasó 17 minutos.

Sea a una función numérica e i un entero positivo. Usamos $S^{-i}\mathbf{a}$ para denotar una función numérica tal que su valor en r es a_{r+i} para $r \geqq 0$. Por ejemplo, sea

$$a_r = \begin{cases} 1 & 0 \leqq r \leqq 10 \\ 2 & r \geqq 11 \end{cases}$$

Si b es $S^{-7}\mathbf{a}$, entonces

$$b_r = \begin{cases} 1 & 0 \leqq r \leqq 3 \\ 2 & r \geqq 4 \end{cases}$$

Como otro ejemplo, sea a una función numérica donde a_r es el número de aspirantes cuya altura es r pulgadas por arriba de la altura mínima requerida para enrolarse en el ejército. Se tiene que la función numérica $S^{-2}\mathbf{a}$ describe al número de aspirantes de los grupos de estaturas si se han agregado dos pulgadas a la altura mínima requerida.

La *suma acumulada* de una función numérica a es una función numérica cuyo valor en r es igual a $\sum_{i=0}^{r} a_i$. Por ejemplo, si a describe los ingresos mensuales de una empleada, y si b es la suma acumulada de a, entonces b da sus ingresos acumulados por mes. Otro

$$a_r = 100(1.07)^r \qquad r \geqq 0$$

Si b es la suma acumulada de a, entonces

$$b_r = \sum_{i=0}^{r} a_i = \sum_{i=0}^{r} 100(1.07)^i = \frac{10\,000}{7}[(1.07)^{r+1} - 1] \qquad r \geqq 0$$

ejemplo

Si depositamos $100 cada año en una cuenta de ahorros con una tasa anual de interés compuesto del 7 por ciento, entonces b_r es la cantidad total que tenemos en la cuenta después de r años.

La *diferencia hacia adelante* de una función numérica a es una función numérica, denotada por $\Delta\mathbf{a}$, cuyo valor en r es igual $a_{r+1} - a_r$. La *diferencia hacia atrás* de una función numérica a es una función numérica, denotada por Va, cuyo valor es igual a a_0 en 0 y es igual a $a_r - a_{r-1}$ en $r \geqq 1$. Así, si a describe el ingreso total de una compañía en cada mes, $\Delta\mathbf{a}$ describirá el incremento de ingreso del r-ésimo mes hasta el $(r+1)$-ésimo mes, y $\nabla\mathbf{a}$ describirá el incremento de ingreso del r-ésimo mes respecto del $(r$ - l)-ésimo mes. Por ejemplo, sea a una función numérica que

$$a_r = \begin{cases} 0 & 0 \leqq r \leqq 2 \\ 2^{-r} + 5 & r \geqq 3 \end{cases}$$

Si b denota la diferencia hacia adelante de a ,$\Delta \mathbf{a}$; entonces

$$b_r = \begin{cases} 0 & 0 \leqq r \leqq 1 \\ \frac{41}{8} & r = 2 \\ -2^{-(r+1)} & r \geqq 3 \end{cases}$$

Si c denota la diferencia hacia atrás de a, $\nabla \mathbf{a}$; entonces

$$c_r = \begin{cases} 0 & 0 \leqq r \leqq 1 \\ \frac{41}{8} & r = 2 \\ -2^{-r} & r \geqq 3 \end{cases}$$

Observemos que $S^{-1}(\nabla \mathbf{a}) = \Delta \mathbf{a}$.

Sean a y b dos funciones numéricas. La *convolución* de a y b, denotada por a * b, es una función numérica c que

$$c_r = a_0b_r + a_1b_{r-1} + a_2b_{r-2} + \cdots + a_ib_{r-i} + \cdots + a_{r-1}b_1 + a_rb_0 = \sum_{i=0}^{r} a_ib_{r-i}$$

Por ejemplo, sean a y b dos funciones numéricas tales que

$$a_r = 3^r \qquad r \geqq 0$$

y

$$b_r = 2^r \qquad r \geqq 0$$

Entonces, para c = a * b

$$c_r = \sum_{i=0}^{r} 3^i 2^{r-i} \qquad r \geqq 0$$

Ejemplo 9-3

Consideremos el problema de determinar c_n el número de sucesiones de longitud r que se construyen con las letras$\{x, y, z, \alpha, \beta\}$, con la primera porción de cada sucesión formada con letras del castellano y la segunda porción con letras griegas. Puesto que

$$a_r = 3^r$$

es el número de sucesiones de longitud r formadas con las letras $\{x, y, z\}$ y

$$b_r = 2^r$$

es el número de sucesiones de longitud r formadas con las letras $\{\alpha, \beta\}$,

$$c_r = \sum_{i=0}^{r} 3^i 2^{r-i}$$

es la respuesta a nuestra pregunta.

❐

Ejemplo 9-4

De nuevo consideremos el ejemplo de la cuenta de ahorros que recibe intereses compuestos a una tasa del 7 por ciento anual. Supongamos que al inicio depositamos

\$100 en la cuenta, \$110 al final del primer año, \$120 al final del segundo año, \$100(1 + 0.1*r)* al final del *r*-ésimo año, y así sucesivamente. Queremos saber la cantidad que hay en la cuenta al final del *r*-ésimo año. Si la función numérica a describe el depósito anual, esto es,

$$a_r = 100(1 + 0.1r) \qquad r \geqq 0$$

Si b es una función numérica tal que

$$b_r = (1.07)^r \qquad r \geqq 0$$

Observamos que b_r es la cantidad en la cuenta de ahorros al final del *r*-ésimo año si se deposita \$1 en la cuenta al inicio (o al final del 0-ésimo año). Así, si a_i dólares se depositaron en la cuenta de ahorros al final del *i*-ésimo año, éstos se convertirán en $a_i b_{r-i}$ dólares en *r* - *i* años después (o al final del *r*-ésimo año). Se tiene que

$$c_r = \sum_{i=0}^{r} a_i b_{r-i}$$

es la cantidad total en la cuenta de ahorros al final del *r*-ésimo año.

❐

Ejemplo 9.5

Consideremos la operación financiera de una compañía. Por cada \$100 000 en órdenes que recibe, la compañía debe tomar prestados \$70 000 como capital activo para materiales, maquinaria, salarios, etcétera. En específico, la compañía necesitará \$30 000 de inmediato, \$20 000 en el siguiente mes y \$5 000 en cada uno de los cuatro meses siguientes. Así, las necesidades de nuevo capital activo, en miles de dólares, pueden ser descritas mediante una función numérica a tal que

$$a_r = \begin{cases} 30 & r = 0 \\ 20 & r = 1 \\ 5 & 2 \leqq r \leqq 5 \\ 0 & r \geqq 6 \end{cases}$$

Por cada \$ 1000 que la compañía toma prestados, debe liquidar un total de \$ 1100 en 11 pagos de \$100 para cada uno de los 11 meses a partir del mes siguiente a aquel en que se tomó el préstamo. Sea b una función numérica que describe el calendario de pagos, en miles de dólares. Tenemos que

$$b_r = \begin{cases} 0 & r = 0 \\ 0.1 & r = 1, 2, \ldots, 11 \\ 0 & r \geqq 12 \end{cases}$$

Supongamos que la compañía tiene en órdenes un total de \$200 000 en cada uno de los 10 primeros meses y en los meses siguientes un total en órdenes de \$300 000. Queremos saber el total de pago en el *r*-ésimo mes. Sea c una función numérica tal que

$$c_r = \begin{cases} 2 & 0 \leqq r \leqq 9 \\ 3 & r \geqq 10 \end{cases}$$

representa las órdenes recibidas en el r-ésimo mes, en cientos de miles de dólares. Observemos que **a * b** dará el pago mensual correspondiente a una sola orden por $ 100 000 al inicio. Exhortamos al lector a verificar que a * b = (0, 3, 5, 5.5, 6, 6.5, 7, 7, 7, 7, 7, 7,4,2,1.5,1, .5,0,0,...). En consecuencia, c * (a * b) dará el calendario de pagos de todas las deudas en miles de dólares.* Exhortamos al lector a verificar que c * (a * b) = (0, 6, 16, 21,...)

9.3 COMPORTAMIENTO ASINTÓTICO DE LAS FUNCIONES NUMÉRICAS

Sea a una función numérica que describe el valor en el mercado de una propiedad de tierra de cultivo que hemos adquirido. En particular, sea a_r el valor de la propiedad después de r años. Si compramos la propiedad como una inversión a largo plazo, deberíamos estar más interesados en cómo se incrementa o decrece a_r para un r muy grande. Sea b una función numérica en la cual b_r es el tiempo que le toma a una computadora procesar la nómina de r empleados. Para asegurar la eficiencia de la operación de procesamiento de datos, estamos más interesados en saber la cantidad de tiempo que toma procesar las nóminas con un número grande de empleados, porque si se tiene un número pequeño de empleados, el tiempo de procesamiento total no será prohibitivo bajo ninguna circunstancia y, en consecuencia, no es un problema serio. Sea c una función numérica en la cual c_r es el tiempo que le toma a cierto algoritmo de ordenamiento ordenar r números. Estamos interesados en cómo se incrementa c_r cuando queremos usar el algoritmo para ordenar decenas de miles de nombres durante la preparación de un directorio telefónico. Por *comportamiento asintótico* de una función numérica, entendemos el modo en que varía el valor de la función para r muy grande. Por ejemplo, para

$$a_r = 5 \qquad r \geqq 0$$

el valor de la función numérica permanece constante conforme se incrementa r. Para

$$b_r = 5r^2 \qquad r \geqq 0$$

el valor de la función se incrementa conforme se incrementa r, y es proporcional a r^2, en tanto que para

$$c_r = 5 \log r \qquad r \geqq 0$$

el valor de la función se incrementa conforme se incrementa r, y es proporcional a log r. Por último, para

$$d_r = \frac{5}{r} \qquad r \geqq 0$$

el valor de la función decrece conforme se incrementa r, es proporcional a $1/r$, y se aproxima a 0 como un límite.

† Exhortamos al lector a verificar que si expresamos a_r y b_r en miles de dólares y c_r en cientos de miles de dólares obtendremos el calendario de pagos en miles de dólares.

En muchas ocasiones estamos interesados en comparar el comportamiento asintótico de dos funciones numéricas. Para esto, introducimos la noción de *dominio asintótico.* En las funciones numéricas **a** y **b**, decimos que a *domina asintóticamente* a **b** o **b** es *asintóticamente dominada* por **a**, si existen las constantes positivas k y m, tales que†

$$|b_r| \leqq ma_r \qquad \text{para } r \geqq k$$

Por ejemplo, sean **a** y **b** dos funciones numéricas que

$$a_r = r + 1 \qquad r \geqq 0$$
$$b_r = \frac{1}{r} + 7 \qquad r \geqq 0$$

Entonces **a** domina asintóticamente a **b** porque, para $k = 7$ y $m = 1$,

$$|b_r| \leqq a_r \qquad \text{para } r \geqq 7$$

Por otro lado, b no domina asintóticamente a **a**, ya que para cualquier elección de k y m, existe un r_0 tal que $r_0 \geqq k$ y $|a_{r0}| > mb_{r0}$. Para otro ejemplo, sean **a** y **b** dos funciones numéricas tales que

$$a_r = \frac{1}{100}r^2 - 1000$$
$$b_r = \begin{cases} -1\,000\,000 & 0 \leqq r \leqq 10 \\ -100\,000r & r > 10 \end{cases}$$

Debido a que para $k = 10^6 + 1$ y $m = 10$,

$$|b_r| \leqq 10a_r \qquad \text{para } r \geqq 10^6 + 1$$

podemos concluir que **a** domina asintóticamente a **b**.

Si **a** domina asintóticamente a **b**, podemos intuir que **a** crece más rápido que **b**. Así, para una r suficientemente grande, el valor absoluto de b_r no excede una proporción fija de a_r. Por ejemplo, supongamos un depósito de 1 dólar en dos cuentas bancarias separadas. Después de r años, nuestro depósito total en la cuenta A está dado por 1 + 0.2r dólares, y nuestro depósito total en la cuenta B es de $(1 + 0.03)^{4r}$ dólares.‡ Si

$$a_r = 1 + 0.2r \qquad r \geqq 0$$
$$b_r = (1 + 0.03)^{4r} \qquad r \geqq 0$$

Observamos que b domina asintóticamente a a puesto que, para $k = 9$ y $m = 1$,

$$|1 + 0.2r| \leqq (1 + 0.03)^{4r} \qquad \text{para } r \geqq 9$$

Esto es, la razón de crecimiento de nuestro dinero en la cuenta B es mayor que la de la cuenta A. Aunque durante algunos de los primeros años el total en la cuenta B es menor que el de la cuenta A, el total en la cuenta B supera al de la cuenta A a largo plazo.

---

† Algunos autores utilizan la condición $|b_r| \leqq m|a_r|$ para $r \geqq k$ en lugar de la que hemos dado aquí. No obstante, como verá el lector más adelante, es muy significativo el insistir en que a_r sea positivo.

‡ Esto es, la cuenta .4 paga el 20 por ciento anual de interés simple (una práctica bancada en cierta forma antigua), y la cuenta B paga el 12 por ciento anual de interés compuesto cada cuatro meses.

Dejamos las demostraciones de las siguientes proposiciones al lector:

1. Para cualquier función numérica a, |**a**| domina asintóticamente a **a**.
2. Si **b** es dominada asintóticamente por **a**, entonces para toda constante α, αb también es dominada asintóticamente por **a**.
3. Si **b** es dominada asintóticamente por a, entonces para todo entero i, $S^i\mathbf{b}$ es dominada asintóticamente por $S^i\mathbf{a}$.
4. Si tanto **b** como **c** son dominadas asintóticamente por **a**, entonces las constantes arbitrarias α y β, $\alpha\mathbf{b} + \beta\mathbf{c}$ también es dominada asintóticamente por **a**.
5. Si **c** es dominada asintóticamente por **b**, y **b** es dominada asintóticamente por **a**, entonces **c** es dominada asintóticamente por **a**.
6. Es posible que **a** sea dominada asintóticamente por **b**, y que **b** sea dominada asintóticamente por **a**. Por ejemplo, si

$$a_r = r^2 + r + 1 \qquad r \geqq 0$$
$$b_r = 0.05r^2 - r^{1/3} - 9 \qquad r \geqq 0$$

7. Es posible que **a** no domine asintóticamente **a b**, y que **b** tampoco domine asintóticamente a **a**. Por ejemplo, si

$$a_r = \begin{cases} 1 & r \text{ es par} \\ 0 & r \text{ es impar} \end{cases}$$

$$b_r = \begin{cases} 0 & r \text{ es par} \\ 1 & r \text{ es impar} \end{cases}$$

8. Es posible que tanto **a** como **b** dominen asintóticamente a **c**, en tanto que a no domina asintóticamente **a b**, y **b** tampoco domina asintóticamente a **a**. Por ejemplo, si

$$a_r = \begin{cases} 1 & r = 3i \text{ o } 3i + 1 \\ 0 & r = 3i + 2 \end{cases}$$

$$b_r = \begin{cases} 1 & r = 3i \text{ o } 3i + 2 \\ 0 & r = 3i + 1 \end{cases}$$

$$c_r = \begin{cases} 1 & r = 3i \\ 0 & \text{en otro caso} \end{cases}$$

Para una función numérica discreta dada a, denotaremos por $O(\mathbf{a})$ al conjunto de *todas* las funciones numéricas que son dominadas asintóticamente por **a**. $O(\mathbf{a})$ se lee como "orden **a**" o como "o-grande de a". Así, si b es dominada asintóticamente por a, entonces **b** está en el conjunto $O(\mathbf{a})$. En lugar de decir que **b** está en el conjunto de con $O(\mathbf{a})$, frecuencia diremos que **b** es $O(\mathbf{a})$. Por ejemplo,

$$\mathbf{a} = r^2$$
$$\mathbf{b} = 10r^2 + 25$$
$$\mathbf{c} = 5 - r$$
$$\mathbf{d} = \frac{1}{2} r^2 \log r - r^2$$

Notamos que **b** es $O(\mathbf{a})$, **c** es $O(\mathbf{a})$ pero **d** no es $O(\mathbf{a})$. También **a** es $O(\mathbf{d})$, como lo son **b** y **c**.

Cuando una función numérica a puede ser expresada en una forma cerrada simple, como $\mathbf{a} = r^2$, también escribiremos $O(\mathbf{a})$ como $O(r^2)$. Así, cuando decimos que una función numérica $\mathbf{b}$ es $O(r \log r)$, significa que b es asintóticamente dominada por la función numérica a = r log r. Queremos recordar al lector que si $\mathbf{b}$ es $O(\mathbf{a})$, sólo significa que b *no* crece más rápido que a, en realidad, b puede crecer mucho más lentamente que a. Así, una proposición como "b crece demasiado rápido porque $\mathbf{b}$ es $O(2^r)$" no es válida porque sólo sabemos que b no crece más rápido que 2^r. Es posible que b crezca mucho más lentamente que 2^r.

Observemos que las relaciones establecidas con anterioridad, en términos de dominio asintótico, también pueden reestablecerse en términos de las notaciones de "o-grande" de la siguiente manera:

1. Para toda función numérica a, **a** es $O(|\mathbf{a}|)$.
2. Si **b** es $O(\mathbf{a})$, entonces para toda constante **a**, **ab** también es $O(\mathbf{a})$.
3. Si **b** es $O(\mathbf{a})$, entonces para todo entero i, $S^i\mathbf{b}$ es $O(S^i\mathbf{a})$.
4. Si tanto **b** como **c** son $O(\mathbf{a})$, entonces para cualesquiera constantes α y β arbitrarias, $\alpha\mathbf{b} + \beta\mathbf{c}$ también es $O(\mathbf{a})$.
5. Si **c** es $O(\mathbf{b})$ y **b** es $O(\mathbf{a})$, entonces **c** es $O(\mathbf{a})$.
6. Es posible que **a** sea $O(\mathbf{b})$, y **b** sea $O(\mathbf{a})$.
7. Es posible que **a** no sea $O(\mathbf{b})$ y **b** no sea $O(\mathbf{a})$.
8. Es posible que **c** sea tanto $O(\mathbf{a})$ como $O(\mathbf{b})$, en tanto **a** no sea $O(\mathbf{b})$ y **b** no sea $O(\mathbf{a})$.

Consideremos los siguientes ejemplos:

Ejemplo 9-6

Mostremos que para $i \leqq j$, αr^i es $O(r^j)$. Observamos que para $k = 1$ y $m = 1$,

$$|r^i| \leqq r^j \qquad \text{para } r \geqq 1$$

De esto se obtiene que r^i es $O(r^j)$. En consecuencia, αr^i es $O(r^j)$. ❒

Ejemplo 9.7

Sea

$$\mathbf{a} = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \cdots + \alpha_n r^n$$

Queremos demostrar que $\mathbf{a}$ es $O(r^n)$. Observamos que

$$\begin{aligned} |\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \cdots + \alpha_n r^n| &\leqq |\alpha_0| + |\alpha_1| r + |\alpha_2| r^2 + \cdots + |\alpha_n| r^n \\ &\leqq (|\alpha_0| + |\alpha_1| + |\alpha_2| + \cdots + |\alpha_n|) r^n \end{aligned}$$

Puesto que para $k = 1$ y $m = |\alpha_0| + |\alpha_1| + |\alpha_2| + \cdots + |\alpha_n|$

$$|\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \cdots + \alpha_n r^n| \leqq m r^n \qquad \text{para } r \geq 1$$

a es dominada asintóticamente por r^n. Por tanto, a es $O(r^n)$.

De hecho, el resultado también puede obtenerse más directamente. De acuerdo con el ejemplo 9.6, para $i \leqq n$, $\alpha_i r^i$ es $O(r^n)$. Si usamos el resultado en la relación 4 anterior, podemos concluir que a es $O(r^n)$. ❒

Ejemplo 9-8

Observemos que $O(1) \subset O(\log r) \subset O(r) \subset O(r^i) \subset O(\alpha^r) \subset O(r!)$. Exhortamos al lector a confirmar los resultados. ❐

Sean A y B dos conjuntos de funciones numéricas. Usamos las siguientes notaciones para denotar conjuntos de funciones numéricas:

$$\mathbf{A} + \mathbf{B} = \{\mathbf{a} + \mathbf{b} | \mathbf{a} \in \mathbf{A}, \mathbf{b} \in \mathbf{B}\}$$
$$\alpha\mathbf{A} = \{\alpha\mathbf{a} | \mathbf{a} \in \mathbf{A}\}$$
$$\mathbf{A} \cdot \mathbf{B} = \{\mathbf{ab} | \mathbf{a} \in \mathbf{A}, \mathbf{b} \in \mathbf{B}\}$$

Establecemos los siguientes resultados y dejamos las demostraciones al lector:

1. Si $\mathbf{b}$ es $O(\mathbf{a})$, entonces $O(\mathbf{b})$ es un subconjunto de $O(\mathbf{a})$. Por tanto, si $\mathbf{b}$ es $O(\mathbf{a})$ y $\mathbf{a}$ es $O(\mathbf{b})$, entonces los conjuntos $O(\mathbf{a})$ y $O(\mathbf{b})$ son iguales.
2. Para toda $\mathbf{a}$, $O(\mathbf{a}) + O(\mathbf{a}) = O(\mathbf{a})$.
3. Si $\mathbf{b} \in O(\mathbf{a})$, entonces $O(\mathbf{a}) + O(\mathbf{b}) = O(\mathbf{a})$.
4. Para cualquier constante α, $\alpha O(\mathbf{a}) = O(\alpha\mathbf{a}) = O(\mathbf{a})$.
5. Para cualquier $\mathbf{a}$ y $\mathbf{b}$, $O(\mathbf{a})O(\mathbf{b}) = O(\mathbf{ab})$.

Consideremos los siguientes ejemplos:

Ejemplo 9-9

Sea

$$\mathbf{a} = \tfrac{1}{3}r^3 + \tfrac{1}{2}r^2 + r$$

Es obvio, a es $O(r^3)$. Además, a está en el conjunto de funciones numéricas

$$\{\tfrac{1}{3}r^3\} + O(r^2)^\dagger$$

también como en el conjunto de funciones numéricas

$$\{\tfrac{1}{3}r^3 + \tfrac{1}{2}r^2\} + O(r)$$

En la literatura, con frecuencia escribimos

$$\mathbf{a} = O(r^3) \tag{9.1}$$
$$\mathbf{a} = \tfrac{1}{3}r^3 + O(r^2) \tag{9.2}$$
$$\mathbf{a} = \tfrac{1}{3}r^3 + \tfrac{1}{2}r^2 + O(r) \tag{9.3}$$

para indicar que‡

$$\{\mathbf{a}\} \in O(r^3)$$
$$\{\mathbf{a}\} \in \{\tfrac{1}{3}r^3\} + O(r^2)$$
$$\{\mathbf{a}\} \in \{\tfrac{1}{3}r^3 + \tfrac{1}{2}r^2\} + O(r)$$

---

† Le recordamos al lector que esto denota el conjunto de funciones numéricas A + B, donde A es el conjunto $\{\tfrac{1}{3}r^3\}$ y B es el conjunto $O(r^2)$.

‡ Algunos autores prefieren la interpretación:

$$\{\mathbf{a}\} \subseteq O(r^3)$$
$$\{\mathbf{a}\} \subseteq \{\tfrac{1}{3}r^3\} + O(r^2)$$
$$\{\mathbf{a}\} \subseteq \{\tfrac{1}{3}r^3 + \tfrac{1}{2}r^2\} + O(r)$$

Estas ligeras diferencias en el punto de vista no tienen importantes consecuencias.

Observemos que (9.1), (9.2) y (9.3) proporcionan información diferente sobre la función numérica a. Debido a que

$$\{\tfrac{1}{3}r^3 + \tfrac{1}{2}r^2\} + O(r) \subset \{\tfrac{1}{3}r^3\} + O(r^2) \subset O(r^3)$$

(9.3) es la más fuerte y (9.1) la más débil de las tres relaciones en lo referente al comportamiento de la función numérica a. En otras palabras, (9.3) implica (9.2) la cual, a su vez, implica (9.1), pero no la inversa. ☐

Ejemplo 9.10

Sean

$$\mathbf{a} = r + O\left(\frac{1}{r}\right)$$

$$\mathbf{b} = \sqrt{r} + O\left(\frac{1}{\sqrt{r}}\right)$$

Tenemos que

$$\begin{aligned}\mathbf{ab} &= \left(r + O\left(\frac{1}{r}\right)\right)\left(\sqrt{r} + O\left(\frac{1}{\sqrt{r}}\right)\right)^{\dagger}\\ &= r^{3/2} + rO\left(\frac{1}{\sqrt{r}}\right) + \sqrt{r}\,O\left(\frac{1}{r}\right) + O\left(\frac{1}{r}\right)O\left(\frac{1}{\sqrt{r}}\right)\\ &= r^{3/2} + O(\sqrt{r}) + O\left(\frac{1}{\sqrt{r}}\right) + O\left(\frac{1}{r^{3/2}}\right)\\ &= r^{3/2} + O(\sqrt{r})\end{aligned}$$

☐

De modo similar a la notación "o-grande", también introduciremos las notaciones "omega-grande" y "theta-grande". Para una función numérica dada **a**, Ω (**a**) denota el conjunto de todas las funciones numéricas **b** tales que existen constantes positivas *ky m* con‡

$$|b_r| \geqq ma_r \qquad \text{para } r \geqq k$$

En otras palabras, si **b** está en $\Omega(\mathbf{a})$, entonces **b** crece al menos tan rápido como **a**. Por ejemplo, sean

$$\begin{aligned} a_r &= 2r + 5 & r &\geqq 0\\ b_r &= r - 2 & r &\geqq 0\\ c_r &= r\log r & r &\geqq 0\\ d_r &= r^{3/2} & r &\geqq 0\end{aligned}$$

Tanto **b** como **c** están en $\Omega(\mathbf{a})$. No obstante, ni **b** ni **c** están en $\Omega(\mathbf{d})$. En la literatura, en lugar de decir que **b** está en $\Omega(\mathbf{a})$, decimos que **b** es $\Omega(\mathbf{a})$.

†Le recordamos al lector que esto denota al conjunto de funciones numéricas A B, donde A es el conjunto $r + O(1/r)$ y **B** es el conjunto $\sqrt{r} + O(1/\sqrt{r})$.

‡La notación es significativa sólo si a_r es positiva.

Para una función numérica a, $\Theta\mathbf{a}$ denota al conjunto de todas las funciones numéricas b tales que existen constantes positivas m, m' y k con†

$$ma_r \leqq |b_r| \leqq m'a_r \qquad \text{para } r \geqq k$$

En otras palabras, si **b** está en $\Theta\mathbf{a}$, entonces **b** crece al mismo ritmo que **a**. Así, en lugar de decir que **b** está en $\Theta(\mathbf{a})$, decimos que **b** es $\Theta(\mathbf{a})$. Por ejemplo, de acuerdo con nuestro análisis de la sección 8.2, la complejidad temporal del algoritmo de ordenamiento burbuja es $\Theta(r^2)$. En consecuencia, la complejidad temporal del problema de ordenamiento es $O(r^2)$. Por otro lado, la complejidad temporal del problema de seleccionar al mayor de r números, y la complejidad temporal del problema de seleccionar al mayor y menor de r números son ambas $\Theta(r)$.

9.4 FUNCIONES GENERADORAS

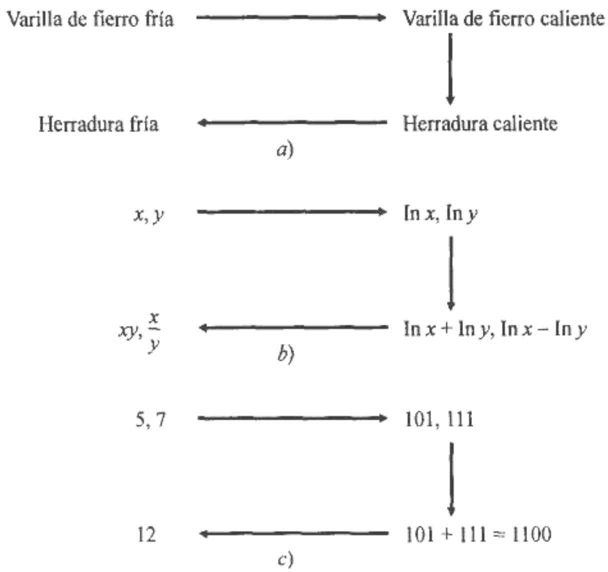
Como señalamos en la sección 9.1, podemos especificar una función numérica mediante un listado exhaustivo de sus valores. En esta sección, introducimos otra forma de representar funciones numéricas.

Supongamos que tenemos una varilla de fierro y con ella deseamos hacer una herradura. Puesto que martillar una varilla de fierro fría hasta lograr una herradura es bastante tedioso, primero colocaremos la varilla de fierro en un horno. Luego podemos martillar la varilla de fierro caliente para obtener una herradura (caliente). Cuando sumerjamos la herradura caliente en un tanque de agua para enfriarla, obtendremos la herradura (fría) que queríamos. La moraleja de nuestro ejemplo es obvia: nuestra meta es convertir una varilla de fierro (fría) en una herradura (fría). Sin embargo, en lugar de alcanzar esta meta en forma directa, primero cambiamos la varilla de fierro fría por una varilla de fierro caliente. Una vez que hacemos una herradura caliente a partir de la varilla de fierro caliente, podemos enfriar la herradura caliente y obtener la herradura fría que queremos. El proceso se muestra en la figura *9.1a.*

Recordemos que a partir de un número positivo x, podemos calcular su logaritmo In x. En efecto, el logaritmo de un número puede ser visto como una representación alternativa del número, ya que a partir de x podemos calcular ln x, y a partir de ln x podemos calcular x. Así, cuando queremos calcular xy, en lugar de multiplicar los dos números x y y directamente, podemos representar x como ln x y y como ln y, calcular ln x + ln y, que es igual a ln xy, y entonces obtener el producto xy a partir de su representación alternativa ln xy. De modo similar, si queremos calcular *xly* en lugar de dividir x por y, podemos calcular ln *{xly)*, que es igual a ln x - ln y, y entonces obtener *xly* a partir de su representación alternativa ln *{xly).* El proceso es delineado en la figura 9.1 A.

Para un estudiante de ciencias de la computación, la noción de representación alternativa es muy conocida. Para un número decimal una representación alternativa es su número correspondiente binario. Así, en lugar de sumar, restar, multiplicar y dividir números decimales

† La notación es significativa sólo si a_r es positiva.



Fisura 9.1

en forma directa, los representamos como números binarios, usamos una computadora para realizar las operaciones aritméticas sobre números binarios (lo que una computadora puede hacer sin esfuerzo), y entonces obtenemos los resultados de nuestros cálculos al convertir los resultados de números binarios a números decimales. El proceso se muestra en la figura 9.1*c*.

Estos ejemplos ilustran muy bien la noción de representación alternativa para entidades físicas, así como matemáticas. Además, vimos que una representación alternativa *escogida adecuadamente* también conduce a la eficiencia y facilidad en algunas de las operaciones que deseamos llevar a cabo.

Ahora introducimos una manera alternativa de representar a las funciones numéricas. Para una función numérica $(a_0, a_1, a_2, \ldots, a_r, \ldots)$, definimos la serie infinita

$$a_0 + a_1z + a_2z^2 + \cdots + a_rz^r + \cdots$$

a la cual llamamos *función generadora* de la función numérica a. En realidad, podemos considerar que hasta el momento sólo hemos introducido una notación nueva . En lugar de escribir cada uno de los valores de una función numérica y usar comas a manera de "separadores" como en $(a_0, a_1, a_2, \ldots, a_r, \ldots)$, escogemos una variable formal *z* y usamos las potencias de *z* como "indicadores" en una serie infinita, tal que el coeficiente de z^r sea el valor de la función numérica en *r.* Para una función numérica **a**, usamos la correspondiente letra mayúscula y escribimos *A(z)* para denotar a la función generadora de **a**. Es claro que a partir de una función numérica podemos obtener con facilidad su función generadora, y viceversa. Por ejemplo, la función generadora de la función numérica $(3^0, 3^1, 3^2, \ldots, 3^r, \ldots)$ es

$$3^0 + 3z + 3^2z^2 + 3^3z^3 + \cdots + 3^rz^r + \cdots \tag{9.4}$$

Observemos que la serie infinita de (9.4) puede escribirse en forma cerrada como

$$\frac{1}{1 - 3z}$$

que es una forma más compacta de representar a la función numérica $(1, 3, 3^2, \ldots, 3^r, \ldots)$. En realidad, como veremos, la razón de introducir la representación de funciones generadoras de funciones numéricas es poder expresar series infinitas en forma cerrada, de manera que puedan manejarse con conveniencia.

El lector puede verificar que $B(z) = \alpha A(z)$, corresponde a $\mathbf{b} = \alpha\mathbf{a}$. Así, por ejemplo, la función generadora de la función numérica

$$a_r = 7 \cdot 3^r \qquad r \geqq 0$$

es

$$A(z) = \frac{7}{1 - 3z}$$

y la función generadora de la función numérica

$$a_r = 3^{r+2} \qquad r \geqq 0$$

es

$$A(z) = \frac{9}{1 - 3z}$$

Es obvio también que $C(z) = A(z) + B(z)$, corresponde a $\mathbf{c} = \mathbf{a} + \mathbf{b}$. Así, la función generadora de la función numérica

$$a_r = 2^r + 3^r \qquad r \geqq 0$$

$$A(z) = \frac{1}{1 - 2z} + \frac{1}{1 - 3z}$$

Del mismo modo, en correspondencia a la función numérica

$$A(z) = \frac{2 + 3z - 6z^2}{1 - 2z}$$

la cual puede escribirse como

$$A(z) = 3z + \frac{2}{1 - 2z}$$

tenemos

$$a_r = \begin{cases} 2 & r = 0 \\ 7 & r = 1 \\ 2^{r+1} & r \geqq 2 \end{cases}$$

Sea a una función numérica y $A(z)$ su función generadora. Sea b una función numérica tal que

$$b_r = \alpha^r a_r$$

para alguna constante α. Puesto que la función generadora de b es

$$a_0 + \alpha a_1 z + \alpha^2 a_2 z^2 + \cdots + \alpha^r a_r z^r + \cdots$$
$$= a_0 + a_1(\alpha z) + a_2(\alpha z)^2 + \cdots + a_r(\alpha z)^r + \cdots$$

tenemos que $B(z) = A(\alpha z)$. Por ejemplo, puesto que la función generadora de la función numérica

$$a_r = 1 \qquad r \geqq 0$$

es

$$A(z) = \frac{1}{1-z}$$

la función generadora de la función numérica

$$a_r = \alpha^r \qquad r \geqq 0$$

es

$$A(z) = \frac{1}{1-\alpha z}$$

Como otro ejemplo, en correspondencia a la función generadora

$$A(z) = \frac{2}{1-4z^2}$$

la cual puede escribirse como

$$A(z) = \frac{1}{1-2z} + \frac{1}{1+2z}$$

tenemos que

$$a_r = 2^r + (-2)^r \qquad r \geqq 0$$

o

$$a_r = \begin{cases} 0 & r \text{ impar} \\ 2^{r+1} & r \text{ par} \end{cases}$$

Resulta desafortunado que para el caso $\mathbf{c} = \mathbf{ab}$, no existe una forma simple con la cual podamos expresar $C(z)$ en términos de $A(z)$ y $B(z)$.

Sea $A(z)$ la función generadora de a. Tenemos que $z^i A(z)$ es la función generadora de $S^i a$ para todo entero positivo /. Por ejemplo, en correspondencia a

$$A(z) = \frac{z^4}{1-2z}$$

tenemos que

$$a_r = \begin{cases} 0 & 0 \leqq r \leqq 3 \\ 2^{r-4} & r \geqq 4 \end{cases}$$

También, $z^{-i}[A(z) - a_0 - a_1 z - a_2 z^2 - \cdots - a_{i-1} z^{i-1}]$ es la función generadora de $S^{-i}\mathbf{a}$. Así, la función generadora de

$$a_r = 3^{r+2} \qquad r \geq 0$$

puede calcularse como

$$\begin{aligned} A(z) &= z^{-2}\left(\frac{1}{1-3z} - 1 - 3z\right) \\ &= z^{-2}\frac{9z^2}{1-3z} \\ &= \frac{9}{1-3z} \end{aligned}$$

Para otro ejemplo, sea

$$a_r = \begin{cases} 0 & 0 \leqq r \leqq 8 \\ 2 & r = 9 \\ 3 & r = 10 \\ 3 & r = 11 \\ 2 & r = 12 \\ 0 & r = 13 \\ 1 & r = 14 \\ 3 & r = 15 \\ 2 & r = 16 \\ 2 & r = 17 \\ 0 & r \geqq 18 \end{cases}$$

la producción por hora de un tornero que llega a trabajar a las 8 a.m., donde r indica la hora del día. Entonces

$$A(z) = z^9(2 + 3z + 3z^2 + 2z^3 + z^5 + 3z^6 + 2z^7 + 2z^8)$$

Supongamos que hay tres torneros. Uno llega a trabajar a las 6 a.m., otro a las 8 a.m. y otro más a las 9 a.m., y desplazan su producción por hora de acuerdo con sus tiempos de inicio. Si consideramos a b como la función numérica que describe la producción total por hora, tenemos que

$$B(z) = (z^{-2} + 1 + z)A(z)$$

Para $\mathbf{b} = \Delta\mathbf{a}$, tenemos

$$B(z) = \frac{1}{z}[A(z) - a_0] - A(z)$$

y para $\mathbf{b} = \nabla\mathbf{a}$, tenemos

$$B(z) = A(z) - zA(z)$$

La representación de una función generadora para funciones numéricas es de particular utilidad al tratar con la convolución de funciones numéricas. Sea $\mathbf{c} = \mathbf{a} * \mathbf{b}$. Puesto que

$$c_r = a_0b_r + a_1b_{r-1} + a_2b_{r-2} + \cdots + a_{r-1}b_1 + a_rb_0$$

es el coeficiente de z^r en el producto

$$(a_0 + a_1z + a_2z^2 + \cdots + a_rz^r + \cdots)(b_0 + b_1z + b_2z^2 + \cdots + b_rz^r + \cdots)$$

Ejemplo 9.11

Sea $\mathbf{c} = \mathbf{a} * \mathbf{b}$ donde

$$a_r = 3^r \qquad r \geqq 0$$

y

$$b_r = 2^r \qquad r \geqq 0$$

Debido a que

$$A(z) = \frac{1}{1-3z}$$

y

$$B(z) = \frac{1}{1-2z}$$

tenemos que

$$C(z) = A(z)B(z) = \frac{1}{1-3z}\,\frac{1}{1-2z} = \frac{3}{1-3z}\,\frac{2}{1-2z}$$

o

$$c_r = 3(3)^r - 2(2)^r = 3^{r+1} - 2^{r+1}$$

❐

concluimos que la función generadora $C(z)$ es igual a $A(z)B(z)$.

Ejemplo 9.12

Sean a una función numérica arbitraria y b la función numérica $(1, 1, 1, \ldots, 1, \ldots)$. Sea

$$\mathbf{c} = \mathbf{a} * \mathbf{b}$$

de manera que

$$c_r = \sum_{i=0}^{r} a_ib_{r-i} = \sum_{i=0}^{r} a_i$$

Así, c es la suma acumulada de a, y la función generadora de c es

$$C(z) = \frac{1}{1-z}\;A(z)$$

Por ejemplo, al dejar que $A(z)$ sea $1/(1-z)$, obtenemos el resultado de que $1/(1-z)^2$ es la función generadora de la función numérica $(1, 2, 3, \ldots, r, \ldots)$.

❐

Ejemplo 9.13 Deseamos evaluar la suma

$$1^2 + 2^2 + 3^2 + \cdots + r^2$$

Primero determinemos la función generadora de la función numérica $(0^2, 1^2, 2^2, 3^2, \ldots, r^2, \ldots)$. Al derivar ambos lados de la igualdad

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + z^4 + \cdots + z^r + \cdots$$

obtenemos

$$\frac{1}{(1-z)^2} = 1 + 2z + 3z^2 + 4z^3 + \cdots + rz^{r-1} + \cdots$$

Se tiene que

$$\frac{d}{dz}\frac{z}{(1-z)^2} = 1^2 + 2^2 \cdot z + 3^2 \cdot z^2 + 4^2 \cdot z^3 + \cdots + r^2 \cdot z^{r-1} + \cdots$$

y que

$$z\frac{d}{dz}\frac{z}{(1-z)^2} = 0^2 + 1^2 \cdot z + 2^2 \cdot z^2 + 3^2 \cdot z^3 + 4^2 \cdot z^4 + \cdots + r^2 \cdot z^r + \cdots$$

Así, $z(d/dz)[z/(1-z)^2]$, que es igual a $[z(1+z)]/(1-z)^3$, es la función generadora de la función numérica $(0^2, 1^2, 2^2, 3^2, \ldots, r^2, \ldots)$, y de aquí se tiene que $[z(1+z)]/(1-z)^4$ es la función generadora de la función numérica $(0^2, 0^2 + 1^2, 0^2 + 1^2 + 2^2, 0^2 + 1^2 + 2^2 + 3^2, \ldots, 0^2 + 1^2 + 2^2 + 3^2 + \cdots + r^2, \ldots)$. De acuerdo con el teorema binomial,†

$$\begin{aligned}\frac{(-4)(-4-1)(-4-2)\cdots(-4-r+1)}{r!}(-1)^r &= \frac{4 \times 5 \times 6 \times \cdots \times (r+3)}{r!}\\ &= \frac{(r+1)(r+2)(r+3)}{1 \cdot 2 \cdot 3}\end{aligned}$$

Por tanto, el coeficiente de z^r en la expansión de $[z(1+z)]/(1-z)^4$ es

$$\frac{r(r+1)(r+2)}{1 \cdot 2 \cdot 3} + \frac{(r-1)r(r+1)}{1 \cdot 2 \cdot 3} = \frac{r(r+1)(2r+1)}{6}$$

esto es,

$$1^2 + 2^2 + 3^2 + \cdots + r^2 = \frac{r(r+1)(2r+1)}{6}$$

❐

---

† El teorema binomial establece

$$(1+z)^n = 1 + \sum_{r=1} \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} z^r$$

donde el límite superior de la sumatoria es n si n es un entero positivo, y es oo en otro caso.

*9.5 PROBLEMAS COMBINATORIOS

La representación de la función generadora para las funciones numéricas es muy útil en la solución de problemas combinatorios. En esta sección presentaremos algunos ejemplos ilustrativos. Consideremos la función numérica a tal que

$$a_r = C(n, r)$$

para un n fijo.† n. La función generadora de a es

$$A(z) = C(n, 0) + C(n, 1)z + C(n, 2)z^2 + \cdots + C(n, r)z^r + \cdots + C(n, n)z^n$$

Consideremos una forma indirecta para seleccionar *r* objetos a partir de *n* objetos. Primero podemos dividir los *n* objetos en dos grupos, uno con *k*objetos y el otro con *n*-*k*objetos, para algún *k* fijo menor que *n*. Entonces podemos seleccionar *i* objetos del primer grupo, con $C(k, i)$ maneras, *yr*-*i* objetos del segundo grupo, con $C(n - k, r - i)$ maneras, para $i = 0, 1, 2, \ldots, r$. Así, tenemos la ecuación

$$C(n, r) = \sum_{i=0}^{r} C(k, i)C(n - k, r - i)$$

por lo que podemos escribir **a = d * e** donde,

$$D(z) = C(k, 0) + C(k, 1)z + C(k, 2)z^2 + \cdots + C(k, k)z^k$$

y

$$E(z) = C(n - k, 0) + C(n - k, 1)z + C(n - k, 2)z^2 + \cdots + C(n - k, n - k)z^{n-k}$$

Si repetimos el argumento un número suficiente de veces, obtenemos

$$A(z) = (1 + z)^n \tag{9.5}$$

debido a que la función generadora de la función numérica (C(l, 0), C(l, 1), 0, 0,...) es 1 + z. Además, señalemos un argumento combinatorio simple que puede usarse para obtener (9.5). Consideremos el coeficiente del término z^r en la expansión de $(1 + z)^n$. Al calcular el producto de los *n* factores 1 + *z*, cada factor "contribuirá" con un 1 o una *z*. Con más especificidad, diremos que al construir el término z^r, *r* de los factores contribuyen en z cada uno y *n* - *r* de los factores contribuyen en 1 cada uno. Por tanto, el coeficiente de z^r es el número de maneras de seleccionar *r* de los *n* 1 + z factores para construir el término z^r, el cual, desde luego, es igual a *C(n, r). A* continuación mostramos algunos resultados que se pueden obtener directamente de (9.5).

Ejemplo 9.14

Si asignamos a z el valor 1 en (9.5), obtenemos

$$C(n, 0) + C(n, 1) + \cdots + C(n, r) + \cdots + C(n, n) = 2^n$$

Esto es, el número de maneras de seleccionar ninguno, uno, dos,..., o *n* objetos a partir de *n* objetos es 2^n.

† Recordemos que $C(n, r) = 1$ para $r = 0$, y $C(n, r) = 0$ para $r > n$.

Al asignar a z el valor -1 en (9.5), obtenemos

$$C(n,0) - C(n,1) + C(n,2) + \cdots + (-1)^r C(n,r) + \cdots$$
$$\cdots + (-1)^n C(n,n) = 0$$

o

$$C(n,0) + C(n,2) + C(n,4) + \cdots = C(n,1) + C(n,3) + C(n,5) + \cdots$$

Esto es, el número de maneras de seleccionar un número par de objetos a partir de *n* objetos es igual al número de maneras de seleccionar un número impar de objetos.
puede demostrarse de diversas maneras. Mediante una manipulación algebraica directa,

Ejemplo 9.15 La relación

$$C(n,r) = C(n-1,r) + C(n-1,r-1) \tag{9.6}$$

obtenemos

$$\begin{aligned} C(n,r) &= \frac{n!}{r!(n-r)!} \\ &= \frac{n!}{r!(n-r)!}\left(\frac{n-r}{n} + \frac{r}{n}\right) \\ &= \frac{(n-1)!}{r!(n-r-1)!} + \frac{(n-1)!}{(r-1)!(n-r)!} \\ &= C(n-1,r) + C(n-1,r-1) \end{aligned}$$

También podemos demostrar (9.6) mediante un argumento combinatorio: si seleccionamos *r* objetos a partir de *n* objetos, existen *C(n* - 1, *r)* maneras de seleccionar *r* objetos para que un objeto particular siempre sea excluido, y $C(n-1, r-1)$ maneras de que ese objeto particular siempre sea incluido. Por tanto, se cumple (9.6). El uso de la función generadora proporciona una tercera demostración de (9.6). Observemos que C(w, *r)* es el coeficiente de z^r en $(1+z)^n$, lo cual puede escribirse como $(1+z)^{n-1} + z(1+z)^{n-1}$. Puesto que el coeficiente de z^r en $(1+z)^{n-1}$ es $C(n-1, r)$, y el coeficiente de x^r en $z(1+z)^{n-1}$ es $C(n-1, r-1)$, se llega de inmediato a (9.6).

Ejemplo 9-16 Ahora presentamos algunos ejemplos más:

Sea $(a_0, a_1, a_2, \ldots)$ una función numérica tal que a_r es igual al número de maneras para seleccionar *r* objetos entre 10 objetos de los cuales uno, que denotaremos por *X,* puede ser seleccionado a lo más dos veces, otro, al que denotaremos como *Y,* puede ser seleccionado a lo más tres veces, y los otros pueden ser seleccionados sólo una vez. Afirmamos que la función generadora de a es

$$A(z) = (1 + z + z^2)(1 + z + z^2 + z^3)(1 + z)^8$$

Observemos que el coeficiente de z^r en $A(z)$ es el número de maneras para construir el término z^r a partir de los factores $1 + z + z^2$, $1 + z + z^2 + z^3$ y los ocho factores $1 + z$. La contribución del factor $1 + z + z^2$ puede ser 1, z o z^2, lo que corresponde a seleccionar al objeto X cero, uno o dos veces. La contribución del factor $1 + z + z^2 + z^3$ puede ser 1 o z, o z^2, o z^3, lo que corresponde a seleccionar el objeto Y cero, uno, dos o tres veces. La contribución de cada uno de los ocho factores $1 + z$ puede ser 1 o z, lo que corresponde a seleccionar cada uno de los ocho objetos restantes cero o una vez. ❐

Ejemplo 9.17

Sea a_r el número de maneras para seleccionar r objetos entre n objetos con repeticiones ilimitadas. Debido a que cada objeto puede ser seleccionado tantas veces como queramos, a_r es igual al coeficiente de z^r en

$$(1 + z + z^2 + \cdots)^n$$

La contribución de cada uno de los factores $1 + z + z^2 + \ldots$ corresponderá al número de veces que uno de los objetos es seleccionado. Debido a que

$$(1 + z + z^2 + \cdots)^n = \left(\frac{1}{1 - z}\right)^n = (1 - z)^{-n}$$

obtenemos

$$\begin{aligned} a_r &= (-1)^r \frac{(-n)(-n-1)\cdots(-n-r+1)}{r!} \\ &= \frac{(n+r-1)!}{r!(n-1)!} \\ &= C(n+r-1, r) \end{aligned}$$

Este resultado ya lo habíamos obtenido en el capítulo 3.

❐

Ejemplo 9.18

Queremos' determinar el número de maneras en las cuales $2t + 1$ canicas pueden distribuirse en tres cajas distintas de modo que ninguna caja contenga más de t canicas. Afirmamos que el coeficiente de z^{2t+1} en

$$A(z) = (1 + z + z^2 + \cdots + z^t)^3$$

será nuestra respuesta. El coeficiente de z^{2t+1} en $A(z)$ es el número de maneras para construir el término z^{2t+1} a partir de tres factores $1 + z + z^2 + \cdots + z^t$. La contribución de cada factor $1 + z + z^2 + \cdots + z^t$ puede ser $1, z, z^2, \ldots,$ o z^t, lo que corresponde a tener ninguna, una, dos, $\ldots$, o t canicas en una caja. Debido a que

$$\begin{aligned} (1 + z + z^2 + \cdots + z^t)^3 &= \left(\frac{1 - z^{t+1}}{1 - z}\right)^3 \\ &= (1 - 3z^{t+1} + 3z^{2t+2} - z^{3t+3})(1 - z)^{-3} \qquad (9.7) \end{aligned}$$

el coeficiente de z^{2t-1} en (9.7) es el coeficiente de z^{2t+1} en $(1-z)^{-3}$ menos el triple del coeficiente de z^t en $(1-z)^{-3}$. Así, esto es

$$C(3+2t+1-1, 2t+1) - 3C(3+t-1, t)$$

lo cual se simplifica como

$$C(2t+3, 2t+1) - 3C(t+2, t)$$

□

9.6 NOTAS Y REFERENCIAS

Véase el capítulo 3 de Beckenbach [1], el capítulo 2 de Liu [6], los capítulos 1 y 5 de Liu y Liu [7], y los capítulos 1 y 2 de Riordan [8] acerca de funciones numéricas y sus funciones generadoras. Sobre el comportamiento asintótico de funciones numéricas, véase el capítulo 1 de Knuth [4] y el capítulo 5 de Stanat y McAllister [9]. Nosotros seguimos la notación introducida por Knuth [5]. Véase el capítulo 6 de Chung [2] y el capítulo 11 de Feller [3] acerca de la aplicación de la técnica de funciones generadoras en teoría de probabilidad.

1. Beckenbach, E. F. (ed.): "Applied Combinatorial Mathematics", John Wiley & Sons, Nueva York, 1964.
2. Chung, K. L.: "Elementary Probability Theory with Stochastic Processes", Springer-Verlag, Nueva York, 1974.
3. Feller, W.: "An Introduction to Probability Theory and Its Applications", Vol. 1, 2ª ed., John Wiley & Sons, Nueva York, 1957.
4. Knuth, D. E.: "The Art of Computer Programming, Vol. 1, Fundamental Algorithms", 2ª ed., Addison-Wesley Publishing Company, Reading, Mass., 1973.
5. Knuth, D. E.: Big Omicron and Big Omega and Big Theta, *SIGACT News*, 8(2): 18-23 (1976).
6. Liu, C. L.: "Introduction to Combinatorial Mathematics", McGraw-Hill Book Company, Nueva York, 1968.
7. Liu, C. L., y J. W. S. Liu: "Linear Systems Analysis", McGraw-Hill Book Company, Nueva York, 1975.
8. Riordan, J.: "An Introduction to Combinatorial Analysis", John Wiley & Sons, Nueva York, 1958.
9. Stanat, D. F, y D. F. McAllister: "Discrete Mathematics ín Computer Science", Prentice-Hall, Englewood Cliffs, N.J., 1977.

PROBLEMAS

9.1 Se deja caer una pelota ping-pong hacia el suelo desde una altura de 20 m. Suponga que la pelota siempre rebota hasta alcanzar la mitad de la altura desde la cual cae.

a) Sea a_r la altura alcanzada en el r-ésimo rebote. Haga un bosquejo de la función numérica **a**.

b) Sea b_r la pérdida de la altura durante el r-ésimo rebote, exprese b_r en términos de a_r. Haga un bosquejo de la función numérica **b**.

c) Una segunda pelota de ping-pong es lanzada desde una altura de 6 m al tiempo que la primera pelota alcanza el punto mas alto de su tercer rebote. Sea c_r la altura que alcanza la segunda pelota en su r-ésimo rebote. Exprese c_r en términos de a_r.

9.2 En un sistema de control de proceso, un dispositivo de monitoreo mide la temperatura al interior de una cámara de reacción química una vez cada 30 s. Sea a_r la r-ésima lectura en grados Celsius. Determine una expresión para a_r, si se sabe que la temperatura aumenta de 100 a 120" a una tasa constante en los primeros 300 s y permanece en 120° de ahí en adelante.

9.3 Sea a una función numérica tal que a_r es igual al residuo cuando el entero r es dividido por 17. Sea b una función numérica tal que b_r es igual a 0 si el entero r es divisible por 3, y es igual a 1 en cualquier otro caso.

á) Sea $c_r = a_r + b_r$. ¿Para qué valores de r es $c_r = 0$? ¿Para qué valores de r es $c_r = 1$?

b) Sea $d_r = a_r b_r$. ¿Para qué valores de r es $d_r = 0$? ¿Para qué valores de r es $d_r = 1$?

9.4 Sea **a** una función numérica tal que

$$a_r = \begin{cases} 2 & 0 \leqq r \leqq 3 \\ 2^{-r+5} & r \geqq 4 \end{cases}$$

a) Determine $S^2\mathbf{a}$ y $S^{-2}\mathbf{a}$.

b) Determine $\Delta\mathbf{a}$ y $\nabla\mathbf{a}$,

9.5 Introducimos la notación

$$\begin{aligned} \Delta^2\mathbf{a} &= \Delta(\Delta\mathbf{a}) \\ \Delta^3\mathbf{a} &= \Delta(\Delta^2\mathbf{a}) \\ &\cdots\cdots\cdots \\ \Delta^i\mathbf{a} &= \Delta(\Delta^{i-1}\mathbf{a}) \end{aligned}$$

a) Sea a una función numérica tal que

$$a_r = r^3 - 2r^2 + 3r + 2$$

Determine $\Delta\mathbf{a}$, $\Delta^2\mathbf{a}$, $\Delta^3\mathbf{a}$, $\Delta^4\mathbf{a}$.

b) Sea **a** una función numérica tal que a_r es un polinomio de la forma

$$\alpha_0 + \alpha_1 r + \alpha_2 r^2 + \cdots + \alpha_k r^k$$

Demuestre que $\Delta^{k+1}\mathbf{a}$ es igual a 0.

9.6 *a)* Sea $\mathbf{c} = \mathbf{ab}$. Demuestre que

$$\Delta c_r = a_{r+1}(\Delta b_r) + b_r(\Delta a_r)$$

b) Sean a y b dos funciones numéricas tales que $a_r = r + 1$ y $b_r = \alpha^r$ para toda $r \geqq 0$. Determine $\Delta(\mathbf{ab})$.

c) Sean a y b dos funciones numéricas. El cociente de a y b denotado por a/b es una función numérica cuyo valor en r es igual a a_r/b_r. Si $\mathbf{d} = \mathbf{a}/\mathbf{b}$, demuestre que

$$\Delta d_r = \frac{b_r(\Delta a_r) - a_r(\Delta b_r)}{b_r b_{r+1}}$$

d) Determine $\Delta(\mathbf{a}/\mathbf{b})$ para las funciones numéricas a y b en el inciso *b)*.

9.7 Sea a una función numérica. Sea $\Delta^{-1}\mathbf{a}$ la suma acumulada de a. ¿Son $\Delta(\Delta^{-1}\mathbf{a})$ y $\Delta^{-1}(\Delta\mathbf{a})$ las mismas funciones numéricas?

9.8 El interés producido en una cuenta de ahorros se paga con una tasa del 0.5 por ciento mensual, con un interés compuesto mensualmente. Suponga que se depositan $50 en una cuenta de ahorros cada mes por un periodo de cinco años. ¿Cuál es la cantidad total en la cuenta cuatro años después del primer depósito?, ¿y veinte años después del primer depósito?

9.9 Determine **a** * **b** de acuerdo con lo siguiente. Obtenga un bosquejo o una expresión en forma cerrada para la función numérica resultante que sea más conveniente.

a)

$$a_r = \begin{cases} 1 & 0 \leqq r \leqq 2 \\ 0 & r \geqq 3 \end{cases}$$

$$b_r = \begin{cases} 1 & 0 \leqq r \leqq 2 \\ 0 & r \geqq 3 \end{cases}$$

b)

$$a_r = 1 \qquad \text{para toda } r$$

$$b_r = \begin{cases} 1 & r = 1 \\ 2 & r = 3 \\ 3 & r = 5 \\ -6 & r = 7 \\ 0 & \text{en otro caso} \end{cases}$$

c)

$$a_r = 2^r \qquad \text{para toda } r$$

$$b_r = \begin{cases} 0 & 0 \leqq r \leqq 2 \\ 2^r & r \geqq 3 \end{cases}$$

9.10 Sean **a**, **b** y **c** funciones numéricas tales que **a * b = c**. Dado que

$$a_r = \begin{cases} 1 & r = 0 \\ 2 & r = 1 \\ 0 & r \geqq 2 \end{cases}$$

$$c_r = \begin{cases} 1 & r = 0 \\ 0 & r \geqq 1 \end{cases}$$

determine **b**.

9.11 Sean

$$a_r = \begin{cases} 1 & r = 0 \\ 3 & r = 1 \\ 2 & r = 2 \\ 0 & r \geqq 3 \end{cases}$$

$$c_r = 5^r \qquad \text{para toda } r$$

Determine **b** de manera que **a * b = c**.

9.12 *a)* Cada partícula dentro de un reactor nuclear se divide en dos partículas cada segundo. Suponga que se inyecta una partícula en el reactor cada segundo a partir de $t = 0$. ¿Cuántas partículas hay en el reactor en el n-ésimo segundo?

b) Cuando una partícula se divide en dos, una de ellas es la partícula original y la otra es la nueva partícula creada. Suponga que una partícula tiene un tiempo de vida de 10 segundos. Esto es, una partícula creada en el segundo 0 desaparecerá después del décimo segundo. Repita el inciso *a)* y suponga que todas las partículas inyectadas son creadas de nuevo.

9.13 Por un costo de \$1000, una estación de radio proporcionará a una tienda 10 minutos de comerciales diariamente durante cinco días consecutivos y luego 2 minutos de comerciales durante los siguientes cinco días consecutivos. Supongamos que cada minuto de comercial en un día en particular producirá una venta total de \$500 en ese día y una venta total de \$500 x 2^{-n} en el n-ésimo día a partir de ese día.

a) Suponga que le son pagados a la estación de radio \$1000 en el día 0. Determine la función numérica a, donde a_r es la venta total en el r-ésimo día.

b) Suponga que le son pagados diariamente a la estación de radio \$2000 durante 10 días consecutivos a partir del día 0. Determine la función numérica **b** donde b_r es la venta total en el r-ésimo día.

9.14 Sean **a, b** y **c** tres funciones numéricas:

$$\mathbf{a} = 3r - 2$$

$$\mathbf{b} = \frac{2}{r} + 7$$

$$\mathbf{c} = r \ln r$$

a) ¿Domina asintóticamente **a** a **b**? ¿Domina asintóticamente **a** a **c**? ¿Domina asintóticamente **b** a **a**? ¿Domina asintóticamente **b** a **c**? ¿Domina asintóticamente **c** a **a**? ¿Domina asintótica mente **c** a **b**?

b) ¿Domina asintóticamente **a** + **b** a **a**? ¿Domina asintóticamente **a** + **b** a **c**?

c) ¿Domina asintóticamente **ab** a **a**? ¿Domina asintóticamente **a** a **ab**? ¿Domina asintóticamente **ab** a **c**? ¿Domina asintóticamente **c** a **ab**?

d) ¿Domina asintóticamente Δ**a** a **b**?

e) ¿Domina asintóticamente la suma acumulada de **a** a **c**?

9.15 Sean

$$\mathbf{a} = 3^r$$
$$\mathbf{b} = 2^r$$

a) ¿Domina asintóticamente **a** a **b**?

b) ¿Domina asintóticamente **b** a **a**?

c) ¿Domina asintóticamente **a * b** a **a**?

d) ¿Domina asintóticamente **a * b** a **b**?

9.16 Sea

$$a_r = \sum_{i=0}^{r} i^2$$

d) Demuestre que **a** es $O(r^3)$.

b) Demuestre que **a** es $(r^3/3) + O(r^2)$.

c) Sea **b** una función numérica tal que **b** es $O(r^3)$. ¿Es necesario que **b** sea $(r^3/3) + O(r^2)$?

d) Sea **b** una función numérica tal que **b** es $(r^3/3) + O(r^2)$. ¿Es necesario que **b** sea $O(r^3)$?

9.17 Simplifique la expresión $[\ln r + O(1/n)][n + O(\sqrt{n})]$.

9.18 Sea

$$\mathbf{a} = \ln r$$

Demuestre que para $\varepsilon > 0$, **a** es $O(r^{\varepsilon})$.

9.19 Dada una función numérica a, sea

$$b_r = \begin{cases} a_r + a_{r/2} + a_{r/4} + \cdots + a_{r/2i} & r = 2^i \\ 0 & r \neq 2^i \end{cases}$$

Si a es $O(\sqrt{r})$, demuestre que b es $O(\sqrt{r}\ \log r)$.

9.20 Sea

$$\begin{aligned} a_r &= \sum_{i=0}^{\lceil \log_{3/2} r\rceil - 1} \log_2\left(\frac{2}{3}\right)^i r \\ &= \log_2 r + \log_2\left(\frac{2}{3}\right) r + \log_2\left(\frac{4}{9}\right) r + \cdots \end{aligned}$$

Demuestre que a_r es $O(\log^2 r)$.

9.21 Después de un cuidadoso análisis, un estudiante concluye que una determinada función numérica **a** es $O(r \log r)$. Al escuchar eso, su compañera de cuarto le recuerda que ella ha olvidado especificar la base del logaritmo. ¿Ha olvidado ella algo realmente importante?

9.22 *a)* La complejidad temporal del algoritmo A es $O(r^2)$, y la del algoritmo B es $O(r^2 \ln r)$. ¿Podemos concluir que el algoritmo A es superior al algoritmo B?

b) La complejidad temporal del algoritmo A es $O(r^2)$, y la del algoritmo B es $\Omega(r^2 \ln r)$. ¿Podemos concluir que el algoritmo A es superior al algoritmo B?

c) Lacomplejidad temporal del algoritmo A es $\Theta(r^2)$, y la del algoritmos es $\Theta(r^2 \ln r)$. ¿Podemos concluir que el algoritmo A es superior al algoritmo B?

9.23 A dos ingenieros se les ha asignado la tarea de diseñar circuitos integrados. Sea a_r el área total del circuito integrado que contiene r transistores. Se encuentra que para el diseño del ingeniero A, a_r es $O(r^4)$, y para el diseño del ingeniero B, a_r es $\Omega(\sqrt{r})$. En consecuencia el gerente concluye que cuando se incrementa el número de transistores en un circuito, el diseño del ingeniero A ocuparía un área grande, y por tanto tal diseño es malo, en tanto que el diseño del ingeniero B ocuparía un área menor y por tanto es bueno. ¿Sabe el gerente de lo que habla?

9.24 Un profesor invita a sus estudiantes a tomar cerveza un viernes por la tarde. Después de unas cuantas tandas, él dice: "La cantidad total de cerveza que van a consumir es mayor_______ del contenido de mi cartera, y es mayor _______ de su sed". Si otorgamos al profesor el privilegio de abusar de las terminologías matemáticas, ¿se podrían llenar los espacios en blanco para darle significado a la declaración?

9.25 Halle una expresión simple para la función generada por cada una de las siguientes funciones numéricas:

a) $1, -2, 3, -4, 5, -6, \ldots$

b) $1, 2/3, 3/9, 4/27, \ldots, (r+1)/3^r, \ldots$

c) $1, 1, 2, 2, 3, 3, 4, 4, \ldots$

d) $0 \times 1, 1 \times 2, 2 \times 3, 3 \times 4, \ldots$

e) $0 \times 5^0, 1 \times 5^1, 2 \times 5^2, \ldots, r \times 5^r, \ldots$

9.26 Determine la función generada por la función numérica a_n donde

$$a_r = \begin{cases} 2^r & \text{si } r \text{ es par} \\ -2^r & \text{si } r \text{ es impar} \end{cases}$$

9.27 Determine la función numérica discreta correspondiente a cada una de las siguientes funciones generadoras:

a) $A(z) = \dfrac{1}{1 - z^3}$

b) $A(z) = (1 + z)^n + (1 - z)^n$

c) $A(z) = \dfrac{(1+z)^2}{(1-z)^4}$

d) $A(z) = \dfrac{1}{5 - 6z + z^2}$

e) $A(z) = \dfrac{z^5}{5 - 6z + z^2}$

f) $A(z) = \dfrac{7z^2}{(1-2z)(1+3z)}$

g) $A(z) = \dfrac{1 + z^2}{4 - 4z - z^2}$

h) $A(z) = \dfrac{1}{(1-z)(1-z^2)(1-z^3)}$

9.28 *a)* Si a_r denota el número de maneras en que se puede obtener la suma r cuando se lanzan dos dados indistinguibles, determine $A(z)$.

b) Si b_r denota el número de maneras en que se puede obtener la suma r cuando se lanza un solo dado una vez y luego se lanzan dos dados indistinguibles una vez, determine $B(z)$.

9.29 Si a_r denota el número de maneras de sentar a 10 estudiantes en una fila de r sillas, de manera que dos estudiantes no pueden ocupar lugares adyacentes, determine la función generadora de la función numérica discreta $(a_0, a_1, a_2, \ldots, a_r, \ldots)$.

9.30 El tiempo de vida de un conejo es exactamente de 10 años. Suponga que en principio hay dos conejos recién nacidos, y que el número de nacimientos de conejos cada año es dos veces al del año anterior. Determine el número de conejos que hay en el r-ésimo año.

9.31 Sea a_r el número de maneras de seleccionar r pelotas de una fuente infinita de pelotas rojas, azules y blancas, con la restricción de que el número de pelotas rojas seleccionadas siempre es un número par. Determine la función generadora de la función numérica **a**.

9.32 ¿De cuántas maneras pueden ser seleccionadas $3r$ pelotas de entre $2r$ pelotas rojas, $2r$ pelotas azules y $2r$ pelotas blancas?

9.33 Sea a_r el número de maneras de seleccionar r pelotas de entre 100 pelotas rojas, 50 azules y 50 blancas con la restricción de que el número de pelotas rojas seleccionadas no es igual a dos, el de pelotas azules no es igual a tres, y de pelotas blancas no es igual a cuatro. Determine la función generadora de la función numérica a.

9.34 Un conjunto de r pelotas es seleccionado a partir de una fuente infinita de pelotas rojas, azules, blancas y amarillas. Una selección debe satisfacer la condición de que el número de pelotas rojas es par y el número de pelotas azules es impar, o bien, que el número de pelotas blancas es par y el número de pelotas amarillas es impar. Sea a_r el número total de tales selecciones.

a) Determine la función generadora $A(z)$.

b) Halle una expresión en forma cerrada para a_r Evalúe la expresión para $r = 23$.

9.35 Sea a_r el número de maneras de dividir r canicas idénticas en cuatro montones distintos, de manera que cada montón tenga un número impar de canicas mayor o igual a tres.

a) Determine $A(z)$.

b) Determine una expresión en forma cerrada para a_r.

9.36 *a)* Determine la suma

$$\binom{n}{1} + 2\binom{n}{2} + \cdots + i\binom{n}{i} + \cdots + n\binom{n}{n}$$

b) Si escribimos todas las combinaciones de n letras $\{1, 2, \ldots, n\}$, esto es, todas las 1-combinaciones, todas las 2-combinaciones, ¿cuántas veces aparece cada letra?

9.37 Determine la suma

$$C(n, 0) + 2C(n, 1) + 2^2C(n, 2) + \cdots + 2^nC(n, n)$$

9.38 Determine la suma

$$\binom{n}{0}\binom{m}{k} + \binom{n}{1}\binom{m}{k-1} + \binom{n}{2}\binom{m}{k-2} + \cdots + \binom{n}{k}\binom{m}{0}$$

donde $k \leqq n$ y $k \leqq m$.

9.39 *a)* Dado

$$A(z) = (1 + z)^{2n} + z(1 + z)^{2n-1} + \cdots + z^i(1 + z)^{2n-i} + \cdots + z^n(1 + z)^n$$

Demuestre que

$$a_r = \begin{cases} \binom{2n+1}{r} & 0 \leqq r \leqq n \\ \binom{2n+1}{r} - \binom{n}{r-n-1} & n + 1 \leqq r \leqq 2n + 1 \\ 0 & \text{en otro caso} \end{cases}$$

b) Evalúe la suma

$$\binom{2n}{n} + \binom{2n-1}{n-1} + \cdots + \binom{2n-i}{n-i} + \cdots + \binom{n}{0}$$

9.40 *a)* Demuestre que

$$\binom{r}{0}^2 + \binom{r}{1}^2 + \binom{r}{2}^2 + \cdots + \binom{r}{i}^2 + \cdots + \binom{r}{r}^2 = \binom{2r}{r}$$

b) Demuestre que la función generadora para a, donde

$$a_r = \binom{2r}{r}$$

es $(1 - 4z)^{-1/2}$.

CAPÍTULO

DIEZ

Relaciones de recurrencia y algoritmos recursivos

10.1 INTRODUCCIÓN

Supongamos que a un amigo le preguntamos la edad de su hija mayor. Él podría informarnos que ella tiene 19 años. O podría informarnos que es 6 años mayor que su segunda hija. Si le preguntamos por la edad de su segunda hija, en vez de informarnos que tiene 13 años, podría informarnos que es 5 años mayor que su tercera hija. También puede informarnos que su tercera hija es 2 años mayor que su único hijo. En cuanto nos informe que su único hijo tiene 6 años, no deberemos tener dificultad para determinar que su tercera hija tiene 8 años, su segunda hija 13 años, y su hija mayor 19 años.

Consideremos otro ejemplo. Ahora preguntamos por las indicaciones para ir de nuestra casa hacia la estación de ferrocarril. Nos informan, "desde la avenida Prospect, vaya hacia el este sobre la calle Green. Después de pasar la biblioteca pública, continúe sobre la avenida Springfield y luego tome la calle Neil hacia el norte. En la estación de autobuses, dé vuelta hacia la derecha en el semáforo sobre la avenida University. En el segundo semáforo, de vuelta hacia la izquierda y verá la estación de ferrocarril". Desde luego, ésta es una descripción perfecta y clara de cómo ir desde nuestra casa hasta la estación de ferrocarril. Sin embargo, existe otra forma de dar esas indicaciones, la cual es simplemente *"vaya a la terminal de autobuses* y dé vuelta hacia la derecha en el semáforo sobre la avenida University. En el segundo semáforo dé vuelta hacia la izquierda y verá la estación de ferrocarril". Observemos que esta instrucción consiste en dos partes: una nos dice de modo explícito cómo ir de la terminal de autobuses a la estación de ferrocarril, y la otra, simple y concisa, hace uso de nuestro conocimiento de cómo llegar desde nuestra casa hasta la terminal de autobuses. Supongamos qué sucede si no estamos seguros de cómo llegar desde nuestra casa hasta la terminal de autobuses. En ese caso se nos indicará *"vaya a la biblioteca pública* y tome la avenida Springfield. Continúe hacia el norte sobre la calle Neil hasta la terminal de autobuses". Sobra decir que, o sabemos cómo llegar de nuestra casa a la biblioteca pública, o deberemos solicitar más indicaciones.

Sin duda el lector se ha dado cuenta de lo que tratamos de decir. En el primer ejemplo, en vez de decirnos llanamente la edad de su hija mayor, nuestro amigo opta por decirnos la edad de su hija mayor en términos de su segunda hija. Después, en vez de decirnos llanamente la edad de ésta, opta por decirnos la edad de ella en términos de la edad de su tercera hija, etcétera. En el segundo ejemplo, en lugar de ser explícitos en cuanto a todos los detalles de alguna instrucción, especificamos la instrucción en términos del conocimiento previo que tenemos. Podemos hacer varias observaciones acerca de estos dos ejemplos. Primero, usar nuestro conocimiento previo puede ser una manera concisa de proporcionar información o instrucciones; por ejemplo, al mandar a alguien hacia la estación de ferrocarril, una gran cantidad de información puede estar implícita en el simple enunciado, "vaya a la terminal de autobuses". Segundo, necesitamos realizar algún trabajo para hacer uso del conocimiento previo que tenemos. En el primer ejemplo, en cierto momento nuestro amigo deberá decirnos directamente la edad de alguno de sus hijos de manera que podamos determinar las edades de los otros niños. En el segundo ejemplo, necesitamos averiguar cómo llegamos a la biblioteca pública antes de que podamos seguir las instrucciones de cómo llegar de ésta a la estación de ferrocarril. Tercero, podemos tratar de referirnos a algún conocimiento previo en pasos sucesivos (la edad del hijo de nuestro amigo, la de su tercera hija, y así sucesivamente, o cómo llegar a la biblioteca pública, cómo llegar a la terminal de autobuses, etcétera). Dicha cadena de referencias sólo puede determinarse una vez que alcanzamos el punto donde sabemos explícitamente qué hacer sin referirnos a otro conocimiento previo.† En este capítulo, aplicaremos lo que acabamos de aprender, en primer término la especificación de funciones numéricas discretas, y la especificación de algoritmos.

10.2 RELACIONES DE RECURRENCIA

Consideremos la función numérica $\mathbf{a} = (3^0, 3^1, 3^2, \ldots, 3^r, \ldots)$, Es claro que la función puede especificarse mediante una expresión general para a_n a saber.

$$a_r = 3^r \qquad r \geqq 0$$

† Un profesor de Estados Unidos ha escrito un artículo en inglés, su lengua natal, y desea publicarlo en una revista francesa. Puesto que él no sabe francés, le solicita a su colega francés, el profesor Bestougeff, el favor de traducir el artículo. Después de que se terminó la traducción, el profesor de Estados Unidos sintió que debía incluir una nota de pie de página en el artículo para agradecer la contribución de su colega. Él escribió en inglés lo siguiente "El autor desea agradecer al profesor Bestougeffpor traducir al francés el presente artículo", y solicitó al profesor Bestougeff que lo tradujera al francés, lo cual el profesor hizo con mucho gusto. Entonces se le ocurrió al profesor de Estados Unidos que también debería agradecer al profesor Bestougeffpor traducir la nota de pie de página por él. De manera que escribió otra nota de pie de página, "El autor desea agradecer al profesor Bestougeffpor traducir la nota de pie de página anterior", y entonces solicitó al profesor Bestougeffla traducción al francés, y copió la traducción francesa *dos veces* como dos notas al pie de página adicionales. ¡Su problema estaba completamente resuelto!

Como se señaló en el capítulo 9, la función también puede especificarse mediante su función generadora, a saber,

$$A(z) = \frac{1}{1-3z}$$

De acuerdo con nuestro análisis de la sección 10.1, observamos que existe otra manera de especificar una función numérica. Puesto que el valor de a_r es el triple del valor de a_{r-1}, para todo r, una vez que conocemos el valor de a_{r-1} podemos calcular el valor de a_r El valor de a_{r-1} puede, a su vez, ser calculado como el triple del valor de a_{r-2}, el cual, otra vez, es igual al triple del valor de a_{r-3}. Al final, necesitamos el valor de a_0, el cual sabemos que es 1. Así, observamos que la relación

$$a_r = 3a_{r-1}$$

con la información de que $a_0 = 1$ también especifica por completo la función numérica a.

Otro ejemplo, consideremos la sucesión de números† conocida como la *sucesión de números de Fibonacci.* La sucesión comienza con los números 1 y 1, y contiene números que son iguales a la suma de sus dos predecesores inmediatos. Una parte de la sucesión es

$$1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

En este caso es bastante difícil obtener por observación una expresión general para el r-ésimo número de la sucesión, expresión que es

$$a_r = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{r+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{r+1}$$

Tampoco es obvio cuál es la función generadora para la función numérica. Dicha función es

$$A(z) = \frac{1}{1-z-z^2}$$

Por otro lado, la sucesión puede describirse mediante la relación

$$a_r = a_{r-1} + a_{r-2}$$

con la información de que $a_0 = 1$ y $a_1 = 1$.

Para una función numérica $(a_0, a_1, a_2, \ldots, a_r, \ldots)$, una ecuación que relaciona a_n para cualquier r, a una o más de las a_i, $i < r$, es llamada una *relación de recurrencia.* Una relación de recurrencia también se conoce como una *ecuación en diferencias,* y estos dos términos se usarán de modo indistinto. En muchos problemas de cálculo discreto, algunas veces es más fácil obtener la especificación de una función numérica en términos de una relación de recurrencia, que obtener una expresión general para el valor de la función numérica en r o una expresión en forma cerrada para su función generadora. Es obvio que de acuerdo con la

†Es obvio que una función numérica puede verse simplemente como una sucesión de números reales y viceversa.

relación de recurrencia, podemos llevar a cabo un cálculo paso por paso para determinar a_r a partir de a_{r-1}, a_{r-2},..., para determinar a_{r+1}, a partir de a_r $a_{r-1\ l}$,..., y así sucesivamente, siempre que el valor de la función en uno o más puntos permita iniciar el cálculo. Estos valores obtenidos por la función se denominan *condiciones de frontera.* En el primer ejemplo presentado, la condición de frontera es $a_0 = 1$, y en el segundo ejemplo, las condiciones de frontera son $a_0 = 1$ y $a_1 = 1$. Así concluimos que una función numérica puede ser descrita mediante una relación de recurrencia junto a un conjunto apropiado de condiciones de frontera. La función numérica también se conoce como *la solución de la relación de recurrencia.*

Otro paso aparte de la determinación de los valores de una función numérica mediante el cálculo paso por paso, de acuerdo con una relación de recurrencia, es obtener a partir de la relación de recurrencia una expresión general para la solución o una expresión en forma cerrada para su función generadora. Resulta lamentable que no se conozca un método general de solución para manejar todas las relaciones de recurrencia. A continuación estudiaremos una clase de relaciones de recurrencia conocida como *relaciones de recurrencia lineales con coeficientes constantes.*

10.3 RELACIONES DE RECURRENCIA LINEALES CON COEFICIENTES CONSTANTES

Una relación de recurrencia de la forma

$$C_0a_r + C_1a_{r-1} + C_2a_{r-2} + \cdots + C_ka_{r-k} = f(r) \tag{10.1}$$

donde las letras C_i son constantes, se denomina *relación de recurrencia lineal con coeficientes constantes.* La relación de recurrencia de (10.1) se conoce como *relación de recurrencia de* Pésimo *orden,* siempre que tanto C_0 como C_k sean distintos de cero. Por ejemplo,

$$2a_r + 3a_{r-1} = 2^r$$

es una relación de recurrencia lineal de primer orden con coeficientes constantes. Del mismo modo, tanto

$$3a_r - 5a_{r-1} + 2a_{r-2} = r^2 + 5 \tag{10.2}$$

como

$$a_r + 7a_{r-2} = 0$$

son relaciones de recurrencia lineales de segundo orden con coeficientes constantes. En este capítulo restringiremos nuestro análisis a relaciones de recurrencia lineales con coeficientes constantes, ya que con frecuencia encontraremos esta clase de relaciones de recurrencia y sabremos cómo manejarlas.

Consideremos la relación de recurrencia de (10.2). Supongamos que nos dan $a_3 = 3$ y $a_4 = 6$, podemos calcular a_5 como

$$a_5 = \frac{-1}{3}\,[-5 \times 6 + 2 \times 3 - (5^2 + 5)] = 18$$

podemos calcular a_6 como

$$a_6 = \frac{-1}{3}[-5 \times 18 + 2 \times 6 - (6^2 + 5)] = \frac{119}{3}$$

y así

$$a_2 = \frac{-1}{2}[3 \times 6 - 5 \times 3 - (4^2 + 5)] = 9$$

$$a_1 = \frac{-1}{2}[3 \times 3 - 5 \times 9 - (3^2 + 5)] = 25$$

$$a_0 = \frac{-1}{2}[3 \times 9 - 5 \times 25 - (2^2 + 5)] = \frac{107}{2}$$

etcétera. Concluimos que (10.2), con los valores $a_3 = 3$ y $a_4 = 6$, especifican por completo a la función numérica discreta a.

En general, para una relación de recurrencia lineal de *i*-ésimo orden con coeficientes constantes como la mostrada en (10.1), si *k* valores consecutivos de la función numérica a, $a_{m-k}, a_{m-k+1}, \ldots, a_{m-1}$, son conocidos para algún *m*, el valor de a_m puede calcularse de acuerdo con (10.1), a saber,

$$a_m = -\frac{1}{C_0}[C_1 a_{m-1} + C_2 a_{m-2} + \cdots + C_k a_{m-k} - f(m)]$$

Además, el valor de a_{m+1} puede calcularse como

$$a_{m+1} = -\frac{1}{C_0}[C_1 a_m + C_2 a_{m-1} + \cdots + C_k a_{m-k+1} - f(m+1)]$$

y los valores para $a_{m+2}, a_{m+3}, \ldots$ pueden calcularse de manera similar. Asimismo, el valor de a_{m-k-1} puede calcularse como

$$a_{m-k-1} = -\frac{1}{C_k}[C_0 a_{m-1} + C_1 a_{m-2} + \cdots + C_{k-1} a_{m-k} - f(m-1)]$$

y el valor de a_{m-k-2} puede calcularse como

$$a_{m-k-2} = -\frac{1}{C_k}[C_0 a_{m-2} + C_1 a_{m-3} + \cdots + C_{k-1} a_{m-k-1} - f(m-2)]$$

Los valores para $a_{m-k-3}, a_{m-k-4}, \ldots$ pueden calcularse en forma similar. En realidad, para una relación de recurrencia lineal de Pésimo orden, los *k* valores a_i consecutivos siempre son suficientes para determinar de manera única a la función numérica a. En otras palabras, los *k* valores a_i consecutivos constituyen un conjunto apropiado de condiciones de frontera. Por otro lado, para una relación de recurrencia lineal de *k*-ésimo orden con coeficientes constantes, menos de *k* valores de la función numérica no serán suficientes para determinar de manera única a la función numérica. Por ejemplo, sea

$$a_r + a_{r-1} + a_{r-2} = 4 \qquad (10.3)$$

Si sabemos que $a_0 = 2$, podemos encontrar muchas funciones numéricas que satisfagan la relación de recurrencia y la condición de frontera dada. Así,

$$2,\ 0,2,2,\ 0,2,\ 2,\ 0,\ 2,\ 2,\ 0,\ldots$$
$$2,2,\ 0,\ 2,2,\ 0,2,2,\ 0,2,2,\ldots$$
$$2,5,-3,2,5,-3,2,5,-3,2,\ldots$$

son todas posibilidades. Además, más de *k* valores de la función numérica podrían hacer imposible la existencia de una función numérica que satisfaga la relación de recurrencia y las condiciones de frontera dadas. Por ejemplo, para la relación de recurrencia en (10.3), si nos dicen que

$$a_0 = 2 \qquad a_1 = 2 \qquad a_2 = 2$$

entonces obviamente a_0, a_1 y a_2 no satisfacen la relación de recurrencia. En consecuencia, ninguna a puede satisfacer (10.3) y las condiciones de frontera.

Los *k* valores no consecutivos de las a_i podrían o no constituir un conjunto apropiado de condiciones de frontera, según sea la relación de recurrencia específica que tengamos. Aquí no estudiaremos el problema de determinar cuando un conjunto de valores constituye un conjunto apropiado de condiciones de frontera, ya que no es un problema significativo. Véase, no obstante, el problema 10.8.

Debemos señalar que si una relación de recurrencia de &-ésimo orden no es una relación de recurrencia lineal con coeficientes constantes, *k* valores consecutivos de la función numérica podrían no especificar en forma única una solución. Por ejemplo, consideremos la relación de recurrencia

$$a_r^2 + a_{r-1} = 5$$

Si $a_0 = 1$, observamos que

$$1, 2, \sqrt{3}, \ldots$$
$$1, 2, -\sqrt{3}, \ldots$$
$$1, -2, \sqrt{7}, \ldots$$

son todas soluciones para la relación de recurrencia que satisfacen la condición de frontera. Restringiremos nuestro análisis a la solución de relaciones de recurrencia lineales con coeficientes constantes. Hacerlo así implica una ventaja significativa. Ya que sabemos que para un conjunto de condiciones de frontera dadas, la solución a la relación de recurrencia lineal con coeficientes constantes es única, siempre y cuando seamos capaces de encontrar la función numérica que satisfaga las condiciones de frontera así como la relación de recurrencia, y ésta será *la* solución que buscamos. Este argumento empieza a aclarar el "misterio" del procedimiento de solución que vamos a presentar. Determinaremos la solución "adivinando" cuál deberá ser. La justificación para adivinar es simplemente que hacerlo funciona: cuando el procedimiento origina una solución para la relación de recurrencia, ésta será la solución correcta.

10.4 SOLUCIONES HOMOGÉNEAS

La solución (total) para una ecuación en diferencia lineal con coeficientes constantes es la suma de dos partes, la *solución homogénea,* que satisface la ecuación en diferencias cuando el lado derecho de la ecuación se hace 0, y la *solución particular,* que satisface la ecuación en diferencias con $f(r)$ en el lado derecho.† En otras palabras, la función numérica discreta que es solución de la ecuación en diferencias es la suma de dos funciones numéricas discretas -una es la solución homogénea y la otra es la solución particular. Sean $\mathbf{a}^{(h)} = (a_0^{(h)}, a_1^{(h)}, \ldots, a_r^{(h)}, \ldots)$ la solución homogénea y $\mathbf{a}^{(p)} = (a_0^{(p)}, a_1^{(p)}, \ldots, a_r^{(p)}, \ldots)$ la solución particular a la ecuación en diferencias. Puesto que

$$C_0 a_r^{(h)} + C_1 a_{r-1}^{(h)} + \cdots + C_k a_{r-k}^{(h)} = 0 \tag{10.4}$$

y

$$C_0 a_r^{(p)} + C_1 a_{r-1}^{(p)} + \cdots + C_k a_{r-k}^{(p)} = f(r)$$

tenemos que

$$C_0(a_r^{(h)} + a_r^{(p)}) + C_1(a_{r-1}^{(h)} + a_{r-1}^{(p)}) + \cdots + C_k(a_{r-k}^{(h)} + a_{r-k}^{(p)}) = f(r)$$

Es obvio, la solución total, $\mathbf{a} = \mathbf{a}^{(h)} + \mathbf{a}^{(p)}$ satisface la ecuación en diferencias.

En este momento, es probable que el lector esté intrigado por nuestro interés en la solución homogénea de la ecuación en diferencias. En primer lugar, parece que la solución particular podría ser la solución que buscamos. En segundo lugar, la solución homogénea ni siquiera satisface la ecuación en diferencias dada [Observemos que el lado derecho de (10.4) es igual a 0]. No obstante, recordemos que muchas funciones numéricas discretas satisfacen la ecuación en diferencias, pero sólo una de ellas satisface simultáneamente las condiciones de frontera. Como veremos más adelante, en la sección 10.6, la solución particular por sí sola, en general, no satisfará las condiciones de frontera, pero podemos ajustar la solución homogénea para que las soluciones totales satisfagan la ecuación en diferencias, así como las condiciones de frontera.

Una solución homogénea para una ecuación en diferencias lineal con coeficientes constantes es de la forma $A\alpha_1^r$, donde α_1 se conoce como una *raíz característica* y A es una constante determinada por las condiciones de frontera. Si sustituimos $A\alpha^r$ por a_r en la ecuación en diferencias en el lado derecho de la ecuación igual a 0, obtenemos

$$C_0 A\alpha^r + C_1 A\alpha^{r-1} + C_2 A\alpha^{r-2} + \cdots + C_k A\alpha^{r-k} = 0$$

Esta ecuación puede simplificarse como

$$C_0\alpha^k + C_1\alpha^{k-1} + C_2\alpha^{k-2} + \cdots + C_k = 0$$

† Para el lector que ya haya estudiado ecuaciones diferenciales, la analogía entre la solución de ecuaciones diferenciales lineales con coeficientes constantes y la de ecuaciones en diferencias lineales con coeficientes constantes debe ser obvia. En realidad, ésta es la razón de que usemos la frase *ecuaciones en diferencias lineales con coeficientes constantes* en lugar de apegarnos todo el tiempo a la frase *relaciones de recurrencia lineales con coeficientes constantes.*

la cual se conoce como la *ecuación característica* de la ecuación en diferencias. Por tanto, si α_1 es una de las raíces de la ecuación característica (ésta es la razón de que a] se llame raíz característica), $A\alpha_1^r$ es una solución homogénea de la ecuación en diferencias.

Una ecuación característica de *k*-ésimo grado tiene *k* raíces características. Supongamos que las raíces de la ecuación característica son todas distintas. En este caso es sencillo verificar que

$$a_r^{(h)} = A_1\alpha_1^r + A_2\alpha_2^r + \cdots + A_k\alpha_k^r$$

también es una solución homogénea de la ecuación en diferencias, donde $\alpha_1, \alpha_2, \ldots, \alpha_k$ son las distintas raíces características y $A_1, A_2, \ldots, A_k$ son constantes que están determinadas por las condiciones de frontera.†

Ejemplo 10.1

Consideremos de nuevo el ejemplo de la sucesión de números de Fibonacci analizado en la sección 10.2. La relación de recurrencia para la sucesión de números de Fibonacci es

$$a_r = a_{r-1} + a_{r-2}$$

La correspondiente ecuación característica es

$$\alpha^2 - \alpha - 1 = 0$$

la cual tiene dos raíces distintas

$$\alpha_1 = \frac{1+\sqrt{5}}{2} \qquad \alpha_2 = \frac{1-\sqrt{5}}{2}$$

De lo cual se obtiene que

$$a_r^{(h)} = A_1\left(\frac{1+\sqrt{5}}{2}\right)^r + A_2\left(\frac{1-\sqrt{5}}{2}\right)^r$$

es una solución homogénea donde las dos constantes A_l y A_2 serán determinadas a partir de las condiciones de frontera $a_0 = 1$ y $a_1 = 1$.

Ahora supongamos que algunas de las raíces de la ecuación característica son raíces múltiples. Sea α_1 una raíz de multiplicidad *m.* Mostraremos que la correspondiente solución homogénea es

$$(A_1r^{m-1} + A_2r^{m-2} + \cdots + A_{m-2}r^2 + A_{m-1}r + A_m)\alpha_1^r$$

donde las constantes A_i serán determinadas por las condiciones de frontera. Es claro que $A_m\alpha_1^r$ es una solución homogénea de la ecuación en diferencias de (10.1). Para mostrar que $A_{m-1}r\alpha_1^r$ también es una solución homogénea, recordemos que α_1 no sólo es una raíz de la ecuación

$$C_0\alpha^r + C_1\alpha^{r-1} + C_2\alpha^{r-2} + \cdots + C_k\alpha^{r-k} = 0 \tag{10.5}$$

† La pregunta sobre la unicidad de la solución será analizada más adelante.

sino que también es una raíz de la derivada de la ecuación (10.5),

$$C_0 r\alpha^{r-1} + C_1(r-1)\alpha^{r-2} + C_2(r-2)\alpha^{r-3} + \cdots + C_k(r-k)\alpha^{r-k-1} = 0 \qquad (10.6)$$

debido a que α_1 es una raíz múltiple de (10.5). Si multiplicamos (10.6) por $A_{m-1}\alpha$ y remplazamos α por α_1, obtenemos

$$C_0 A_{m-1} r\alpha_1^r + C_1 A_{m-1}(r-1)\alpha_1^{r-1} + C_2 A_{m-1}(r-2)\alpha_1^{r-2} + \cdots + C_k A_{m-1}(r-k)\alpha_1^{r-k} = 0$$

lo cual muestra que $A_{m-1} r\alpha_1^r$ es una solución homogénea.

El hecho de que a] satisfaga la segunda, tercera, . . . , $(m-1)$-ésima ecuaciones derivadas de (10.5) nos permite demostrar que $A_{m-2}r^2\alpha_1^r$, $A_{m-3}r^3\alpha_1^r$, . . . , $A_1 r^{m-1}\alpha_1^r$ también son soluciones homogéneas de manera similar.

Ejemplo 10.2

Consideremos la ecuación en diferencias:

$$a_r + 6a_{r-1} + 12a_{r-2} + 8a_{r-3} = 0$$

La ecuación característica es

$$\alpha^3 + 6\alpha^2 + 12\alpha + 8 = 0$$

Así,

$$a_r^{(h)} = (A_1 r^2 + A_2 r + A_3)(-2)^r$$

es una solución homogénea ya que -2 es una raíz característica triple. ❒

Ejemplo 10.3

Consideremos la ecuación en diferencias

$$4a_r - 20a_{r-1} + 17a_{r-2} - 4a_{r-3} = 0$$

La ecuación característica es

$$4\alpha^3 - 20\alpha^2 + 17\alpha - 4 = 0$$

y las raíces características son $\frac{1}{2}, \frac{1}{2}, 4$. En consecuencia, la solución homogénea es

$$a_r^{(h)} = (A_1 r + A_2)(\tfrac{1}{2})^r + A_3(4)^r$$

❒

10.5 SOLUCIONES PARTICULARES

No existe un procedimiento general para determinar la solución particular de una ecuación en diferencias. No obstante, en casos simples, esta solución puede obtenerse mediante el método de inspección. Como se demostrará en los ejemplos de esta sección, primero estableceremos la forma general de la solución particular de acuerdo con la forma de/(r), y luego determinaremos la solución exacta de acuerdo con la ecuación en diferencias dada. Consideremos la ecuación en diferencias

$$a_r + 5a_{r-1} + 6a_{r-2} = 3r^2 \qquad (10.7)$$

Supongamos que la forma general de la solución particular es

$$P_1r^2 + P_2r + P_3 \tag{10.8}$$

donde P_1, P_2 y P_3 son constantes a determinar. Al sustituir la expresión (10.8) en el lado izquierdo de (10.7), obtenemos

$$\begin{aligned} &P_1r^2 + P_2r + P_3 + 5P_1(r-1)^2 + 5P_2(r-1) + 5P_3 \\ &\qquad + 6P_1(r-2)^2 + 6P_2(r-2) + 6P_3 \end{aligned}$$

lo cual se simplifica como

$$12P_1r^2 - (34P_1 - 12P_2)r + (29P_1 - 17P_2 + 12P_3) \tag{10.9}$$

Si comparamos (10.9) con el lado derecho de (10.7), obtenemos las ecuaciones

$$\begin{aligned} 12P_1 &= 3 \\ 34P_1 - 12P_2 &= 0 \\ 29P_1 - 17P_2 + 12P_3 &= 0 \end{aligned}$$

de lo cual se obtiene

$$P_1 = \tfrac{1}{4} \qquad P_2 = \tfrac{17}{24} \qquad P_3 = \tfrac{115}{288}$$

Por tanto, la solución particular es

$$a_r^{(p)} = \tfrac{1}{4}r^2 + \tfrac{17}{24}r + \tfrac{115}{288}$$

En general, cuando $f(r)$ es de la forma de un polinomio de grado t en r

$$F_1r^t + F_2r^{t-1} + \cdots + F_tr + F_{t+1}$$

la solución particular correspondiente será de la forma

$$P_1r^t + P_2r^{t-1} + \cdots + P_tr + P_{t+1}$$

Ejemplo 10.4

Consideremos la ecuación en diferencias

$$a_r + 5a_{r-1} + 6a_{r-2} = 3r^2 - 2r + 1 \tag{10.10}$$

La solución particular es de la forma

$$P_1r^2 + P_2r + P_3 \tag{10.11}$$

Al sustituir (10.11) en (10.10), obtenemos

$$\begin{aligned} &P_1r^2 + P_2r + P_3 + 5P_1(r-1)^2 + 5P_2(r-1) + 5P_3 \\ &\qquad + 6P_1(r-2)^2 + 6P_2(r-2) + 6P_3 = 3r^2 - 2r + 1 \end{aligned}$$

lo cual se simplifica como

$$12P_1r^2 - (34P_1 - 12P_2)r + (29P_1 - 17P_2 + 12P_3) = 3r^2 - 2r + 1 \tag{10.12}$$

Al comparar los dos lados de (10.12), obtenemos las ecuaciones

$$\begin{aligned} 12P_1 \quad &= 3 \\ 34P_1 - 12P_2 \quad &= 2 \\ 29P_1 - 17P_2 + 12P_3 &= 1 \end{aligned}$$

de lo cual se obtiene

$$P_1 = \tfrac{1}{4} \qquad P_2 = \tfrac{13}{24} \qquad P_3 = \tfrac{71}{288}$$

Por tanto, la solución particular es

$$a_r^{(p)} = \tfrac{1}{4}r^2 + \tfrac{13}{24}r + \tfrac{71}{288}$$

❐

Ejemplo 10.5

Consideremos la ecuación en diferencias

$$a_r - 5a_{r-1} + 6a_{r-2} = 1 \tag{10.13}$$

Puesto que $f(r)$ es una constante, la solución particular también será una constante P. Si sustituimos P en (10.13), obtenemos

$$P - 5P + 6P = 1$$

Esto es,

$$2P = 1$$

o

$$a_r^{(p)} = \tfrac{1}{2}$$

❐

Otro ejemplo, consideremos la ecuación en diferencias

$$a_r + 5a_{r-1} + 6a_{r-2} = 42 \cdot 4^r \tag{10.14}$$

Suponemos que la forma general de la solución particular es

$$P4^r \tag{10.15}$$

Si sustituimos (10.15) en el lado izquierdo de (10.14), obtenemos

$$P4^r + 5P4^{r-1} + 6P4^{r-2}$$

lo cual se simplifica como

$$\tfrac{21}{8}P4^r \tag{10.16}$$

Al comparar (10.16) con el lado derecho de (10.14), obtenemos

$$P = 16$$

Por tanto, la solución particular es

$$a_r^{(p)} = 16 \cdot 4^r$$

En general, cuando $f(r)$ es de la forma β^r, la correspondiente solución particular es de la forma $P\beta^r$, si β no es una raíz característica de la ecuación en diferencias. Además, cuando $f(r)$ es de la forma

$$(F_1 r^t + F_2 r^{t-1} + \cdots + F_t r + F_{t+1})\beta^r$$

la solución particular correspondiente es de la forma

$$(P_1 r^t + P_2 r^{t-1} + \cdots + P_t r + P_{t+1})\beta^r$$

si β no es una raíz característica de la ecuación en diferencias. Analicemos el siguiente ejemplo.

Ejemplo 10.6 Consideremos la ecuación en diferencias

$$a_r + a_{r-1} = 3r2^r \tag{10.17}$$

La forma general de la solución particular es

$$(P_1 r + P_2)2^r \tag{10.18}$$

Al sustituir (10.18) en (10.17), obtenemos

$$(P_1 r + P_2)2^r + [P_1(r-1) + P_2]2^{r-1} = 3r2^r$$

lo cual se simplifica como

$$\tfrac{3}{2}P_1 r2^r + (-\tfrac{1}{2}P_1 + \tfrac{3}{2}P_2)2^r = 3r2^r \tag{10.19}$$

Al comparar los dos lados de (10.19), obtenemos las ecuaciones

$$\tfrac{3}{2}P_1 = 3$$

$$P_1 \quad -\tfrac{1}{2}P_1 + \tfrac{3}{2}P_2 = 0 \quad = \tfrac{2}{3}$$

Así,

$$a_r^{(p)} = (2r + \tfrac{2}{3})2^r$$

y la solución particular es ❐

Para el caso en que β sea una
de la forma

$$(F_1 r^t + F_2 r^{t-1} + \cdots + F_t r + F_{t+1})\beta^r$$

la solución particular correspondiente es de la forma

$$r^{m-1}(P_1 r^t + P_2 r^{t-1} + \cdots + P_t r + P_{t+1})\beta^r$$

Examinemos los siguientes ejemplos.

Ejemplo 10.7 Consideremos la ecuación en diferencias

$$a_r - 2a_{r-1} = 3 \cdot 2^r \tag{10.20}$$

Debido a que 2 es una raíz característica (de multiplicidad 1), la forma general de la solución particular es

$$Pr2^r \tag{10.21}$$

Si sustituimos (10.21) en (10.20), obtenemos

$$Pr2^r - 2P(r-1)2^{r-1} = 3 \cdot 2^r$$

esto es,

$$P2^r = 3 \cdot 2^r$$

o

$$P = 3$$

Así, la solución particular es

$$a_r^{(p)} = 3r2^r$$

❐

Ejemplo 10.8

Para la ecuación en diferencias

$$a_r - 4a_{r-1} + 4a_{r-2} = (r+1)2^r \tag{10.22}$$

ya que 2 es una raíz característica doble, la forma general de la solución particular es

$$r^2(P_1 r + P_2)2^r \tag{10.23}$$

Al sustituir (10.23) en (10.22), obtenemos, después de simplificar:

$$6P_1 r2^r = r2^r$$
$$(-6P_1 + 2P_2)2^r = 2^r$$

de lo cual se obtiene

$$P_1 = \tfrac{1}{6} \qquad P_2 = 1$$

Así, la solución particular es

$$a_r^{(p)} = r^2\left(\frac{r}{6} + 1\right)2^r$$

❐

Ejemplo 10.9

Consideremos la ecuación en diferencias

$$a_r = a_{r-1} + 7 \tag{10.24}$$

Puesto que 1 es una raíz característica de la ecuación en diferencias y 7 puede ser escrito como $7 \cdot 1^r$, la forma general de la solución particular es Pr (el lector deberá averiguar qué sucede si suponemos que la forma general de la solución particular es P). Si sustituimos $a_r^{(p)} = Pr$ en (10.24), obtenemos

$$Pr = P(r-1) + 7$$

esto es

$$P = 7$$

❐

Ejemplo 10.10

Para la ecuación en diferencias

$$a_r - 2a_{r-1} + a_{r-2} = 7$$

hacemos $a_r^{(p)} = Pr^2$. Exhortamos al lector a llevar a
que $P = \frac{7}{2}$.

❐

Ejemplo 10.11

Consideremos la ecuación en diferencias

$$a_r - 5a_{r-1} + 6a_{r-2} = 2^r + r$$

La forma general de la solución particular es

$$P_1 r 2^r + P_2 r + P_3$$

(Observemos que 2 es una raíz característica de la ecuación en diferencias). La sustitución y comparación darán lugar a que

$$P_1 = -2 \qquad P_2 = \tfrac{1}{2} \qquad P_3 = \tfrac{7}{4}$$

y

$$a_r^{(p)} = -r2^{r+1} + \tfrac{1}{2}r + \tfrac{7}{4}$$

❐

10.6 SOLUCIONES TOTALES

Ahora debemos combinar la solución homogénea y la solución particular y determinar los coeficientes indeterminados de la solución homogénea. Para una ecuación en diferencias de yfc-ésimo orden, los k coeficientes indeterminados $A_1, A_2, \ldots, A_k$ de la solución homogénea pueden determinarse mediante las condiciones de frontera, $a_{r_0}, a_{r_0+1}, \ldots, a_{r_0+k-1}$, para cualquier r_0. Supongamos que las raíces características de la ecuación en diferencias son todas distintas. La solución total es de la forma

$$a_r = A_1\alpha_1^r + A_2\alpha_2^r + \cdots + A_k\alpha_k^r + p(r)$$

de ecuaciones lineales: n particular. Así, para $r = r_0, r_0 + 1, \ldots, r_0 + k - 1$, tenemos el sistema

$$\begin{aligned}
a_{r_0} &= A_1\alpha_1^{r_0} + A_2\alpha_2^{r_0} + \cdots + A_k\alpha_k^{r_0} + p(r_0) \\
a_{r_0+1} &= A_1\alpha_1^{r_0+1} + A_2\alpha_2^{r_0+1} + \cdots + A_k\alpha_k^{r_0+1} + p(r_0+1) \qquad (10.25)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{r_0+k-1} &= A_1\alpha_1^{r_0+k-1} + A_2\alpha_2^{r_0+k-1} + \cdots + A_k\alpha_k^{r_0+k-1} + p(r_0+k-1)
\end{aligned}$$

Estas k ecuaciones lineales pueden resolverse para $A_1, A_2, \ldots, A_k$. Por ejemplo, para la ecuación en diferencias de (10.14), la solución total es

$$a_r = A_1(-2)^r + A_2(-3)^r + 16 \cdot 4^r$$

Supongamos que nos dan las condiciones de frontera $a_2 = 278$ y $a_3 = 962$. Al sustituir estos valores,

$$278 = 4A_1 + 9A_2 + 256$$
$$962 = -8A_1 - 27A_2 + 1024$$

obtenemos

$$A_1 = 1 \qquad A_2 = 2$$

Así,

$$a_r = (-2)^r + 2(-3)^r + 16 \cdot 4^r$$

es la solución total de la ecuación en diferencias.

¿Cómo podemos asegurar que las soluciones de las k ecuaciones en (10.25) siempre son únicas? Se puede demostrar matemáticamente que éste es el caso.† No obstante, de la sección 10.2, recordemos que demostramos la unicidad de la solución de una relación de recurrencia lineal con coeficientes constantes de k-ésimo orden para cualesquiera condiciones de frontera consistentes de k valores consecutivos $a_{r_0}, a_{r_0+1}, \ldots, a_{r_0+k-1}$. En consecuencia, la unicidad de la solución de la relación de recurrencia garantiza la unicidad de las soluciones de las k ecuaciones en (10.25). Por otro lado, si nos dan el valor de una función numérica en k puntos no necesariamente consecutivos, aunque podemos establecer un sistema de k ecuaciones para los coeficientes indeterminados $A_1, A_2, \ldots, A_k$ similar al de (10.25), ya que la solución de la relación de recurrencia podría no estar especificada en forma única por dichas condiciones de frontera, no siempre será el caso de que podamos encontrar la solución de estas ecuaciones en forma única.

Cuando no todas las raíces características de la ecuación en diferencias son distintas, se puede llevar a cabo una derivación similar a la presentada con anterioridad. De nuevo, los coeficientes indeterminados en la solución homogénea pueden determinarse en forma única mediante el valor de la función numérica en k puntos consecutivos.

10.7 SOLUCIÓN POR EL MÉTODO DE FUNCIONES GENERADORAS

En lugar de encontrar la solución de una ecuación en diferencias mediante una expresión para el valor de una función numérica como hemos hecho hasta el momento, también podemos determinar la función generadora de la función numérica a partir de la ecuación en diferencias. En muchos casos, una vez que se determina la función generadora, una expresión para el valor de la función numérica puede obtenerse fácilmente. Consideremos la relación de recurrencia

$$a_r = 3a_{r-1} + 2 \qquad r \geqq 1 \tag{10.26}$$

con la condición de frontera $a_0 = 1$. Permítasenos señalar que en (10.26) escribimos de manera explícita (por primera vez en este capítulo) que la relación de recurrencia es válida

† Véase, por ejemplo, el capítulo 3 de Liu [5].

sólo para $r \geqq 1$, un hecho que conocíamos implícitamente a lo largo de todo lo anterior. Observemos que para $r = 0$, la relación de recurrencia se convierte en

$$a_0 = 3a_{-1} + 2$$

Puesto que a_{-1} no está definida, no podemos suponer que la relación de recurrencia sea válida para $r = 0$. En general, dada una ecuación en diferencias de ¿-ésimo orden que especifica una función numérica, debemos conocer para qué valores de r es válida la ecuación. Primero, observemos que la ecuación es válida sólo si $r \geqq k$ debido a que, para $r < k$, la ecuación involucrará valores de a_i que no están definidos. Además, en muchos casos una ecuación en diferencias se origina a partir de un problema físico en el cual el valor de a_i es significativo sólo para $i \geqq t$, para algún t mayor o igual a 0. En ese caso, la ecuación en diferencias es válida sólo para $r - k \geqq t$.

Al multiplicar ambos lados de (10.26) por z^r, obtenemos

$$a_r z^r = 3a_{r-1} z^r + 2z^r \qquad r \geqq 1 \tag{10.27}$$

Si sumamos (10.27) para todo r, $r \geqq 1$, obtenemos

$$\sum_{r=1}^{\infty} a_r z^r = 3 \sum_{r=1}^{\infty} a_{r-1} z^r + 2 \sum_{r=1}^{\infty} z^r$$

Observemos que

$$\sum_{r=1}^{\infty} a_r z^r = A(z) - a_0$$

$$\sum_{r=1}^{\infty} a_{r-1} z^r = z \sum_{r=1}^{\infty} a_{r-1} z^{r-1} = zA(z)$$

$$\sum_{r=1}^{\infty} z^r = \frac{z}{1-z}$$

Luego obtenemos que

Esto es,

$$A(z) - a_0 = 3zA(z) + \frac{2z}{1-z}$$

la cual se simplifica como

$$(1 - 3z)A(z) = \frac{2z}{1-z} + 1$$

o

$$(1 - 3z)A(z) = \frac{1+z}{1-z}$$

$$\begin{aligned} A(z) &= \frac{1+z}{(1-3z)(1-z)} \\ &= \frac{2}{1-3z} - \frac{1}{1-z} \end{aligned}$$

Por tanto, tenemos que

$$a_r = 2 \cdot 3^r - 1 \qquad r \geqq 0$$

Ahora estableceremos un procedimiento general para determinar la función generadora de la función numérica a a partir de la ecuación en diferencias

$$C_0 a_r + C_1 a_{r-1} + C_2 a_{r-2} + \cdots + C_k a_{r-k} = f(r)$$

la cual es válida para $r \geqq s$, donde $s \geqq k$. Al multiplicar ambos lados de esta ecuación por z^r y sumar desde $r = s$ hasta $r = \infty$, obtenemos

$$\sum_{r=s}^{\infty} (C_0 a_r + C_1 a_{r-1} + C_2 a_{r-2} + \cdots + C_k a_{r-k}) z^r = \sum_{r=s}^{\infty} f(r) z^r$$

$$\sum_{r=s}^{\infty} C_0 a_r z^r = C_0[A(z) - a_0 - a_1 z - a_2 z^2 - \cdots - a_{s-1} z^{s-1}]$$

$$\sum_{r=s}^{\infty} C_1 a_{r-1} z^r = C_1 z[A(z) - a_0 - a_1 z - a_2 z^2 - \cdots - a_{s-2} z^{s-2}]$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{r=s}^{\infty} C_k a_{r-k} z^r = C_k z^k[A(z) - a_0 - a_1 z - a_2 z^2 - \cdots - a_{s-k-1} z^{s-k-1}]$$

$$\begin{aligned} A(z) = \frac{1}{C_0 + C_1 z + \cdots + C_k z^k} \Bigg[& \sum_{r=s}^{\infty} f(r) z^r \\ & + C_0(a_0 + a_1 z + a_2 z^2 + \cdots + a_{s-1} z^{s-1}) \\ & + C_1 z(a_0 + a_1 z + a_2 z^2 + \cdots + a_{s-2} z^{s-2}) \\ & + \cdots \\ & + C_k z^k(a_0 + a_1 z + a_2 z^2 + \cdots + a_{s-k-1} z^{s-k-1}) \Bigg] \end{aligned}$$

puesto que tenemos

Abordemos más ejemplos ilustrativos:

Ejemplo 10.12

Supongamos que lanzamos una moneda r veces. Existen 2^r posibles sucesiones de resultados. Queremos conocer el número de sucesiones de resultados en los cuales nunca aparecerán caras en lanzamientos sucesivos. Sea a_r el número de dichas sucesiones. Para cada sucesión de r - 1 caras y cruces en la cual no hay caras consecutivas, podemos agregar una cruz para obtener una sucesión de r caras y cruces en la cual no hay caras consecutivas. Para cada sucesión de r - 2 caras y cruces en la que no hay caras consecutivas, podemos agregar una cruz y luego una cara para obtener una sucesión de r caras y cruces en la cual no hay caras consecutivas. Además, éstas abarcan la totalidad de sucesiones de r caras y cruces en las que no hay caras consecutivas. Tenemos así la relación de recurrencia

$$a_r = a_{r-1} + a_{r-2} \tag{10.28}$$

Observemos que $a_1 = 2$ y $a_2 = 3$. El valor de a_0 no tiene significado físico, y una opción razonable para su valor es 0. En este caso, la ecuación en diferencias *no* es válida para $r = 2$, y es válida sólo para $r \geqq 3$. Ahora procederemos a mostrar cómo podemos obtener *A{z)* a partir de (10.28). Si multiplicamos ambos lados de (10.28) por z^r y sumamos desde r = 3 hasta $r = \infty$, obtenemos

$$\sum_{r=3}^{\infty} a_r z^r = \sum_{r=3}^{\infty} a_{r-1} z^r + \sum_{r=3}^{\infty} a_{r-2} z^r$$

Esto es,

$$A(z) - a_2 z^2 - a_1 z - a_0 = z[A(z) - a_1 z - a_0] + z^2[A(z) - a_0]$$

lo cual se simplifica como

$$A(z) = \frac{z^2 + 2z}{1 - z - z^2} = -1 + \frac{1 + z}{1 - z - z^2}$$

Si expresamos *A(z)* como

$$A(z) = -1 + \frac{(5 + 3\sqrt{5})/10}{1 - [(1 + \sqrt{5})/2]z} - \frac{(-5 + 3\sqrt{5})/10}{1 - [(1 - \sqrt{5})/2]z}$$

vemos que

$$a_r = \begin{cases} 0 & r = 0 \\ \dfrac{5 + 3\sqrt{5}}{10}\left(\dfrac{1 + \sqrt{5}}{2}\right)^r - \dfrac{-5 + 3\sqrt{5}}{10}\left(\dfrac{1 - \sqrt{5}}{2}\right)^r & r \geqq 1 \end{cases}$$

Puesto que el valor de a_0 no tiene significado físico, podemos optar por hacer a_0 = 1 en lugar de 0, de manera que la ecuación en diferencias de (10.28) es válida para r 2; 2. En ese caso, podemos multiplicar ambos lados de (10.28) por z^r y sumar desde $r = 2$ hasta r = oo. Entonces obtenemos

$$\sum_{r=2}^{\infty} a_r z^r = \sum_{r=2}^{\infty} a_{r-1} z^r + \sum_{r=2}^{\infty} a_{r-2} z^r$$

Esto es,

$$A(z) - a_1 z - a_0 = z[A(z) - a_0] + z^2 A(z)$$

o

$$A(z) = \frac{1 + z}{1 - z - z^2}$$

de lo cual se obtiene que

$$a_r = \frac{5+3\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{2}\right)^r - \frac{-5+3\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^r \qquad r \geqq 0$$

□

Ejemplo 10.13

Resolvamos la ecuación en diferencias del ejemplo 10.11 mediante la determinación de la función generadora *A{z).* Para la ecuación en diferencias

$$a_r - 5a_{r-1} + 6a_{r-2} = 2^r + r \qquad r \geqq 2$$

con las condiciones de frontera $a_0 = 1$ y $a_1 = 1$,ya que

$$\sum_{r=2}^{\infty} a_r z^r - 5\sum_{r=2}^{\infty} a_{r-1} z^r + 6\sum_{r=2}^{\infty} a_{r-2} z^r = \sum_{r=2}^{\infty} 2^r z^r + \sum_{r=2}^{\infty} r z^r$$

obtenemos

$$A(z) - a_0 - a_1 z - 5z[A(z) - a_0] + 6z^2 A(z) = \frac{4z^2}{1-2z} + z\left[\frac{1}{(1-z)^2} - 1\right]$$

lo cual se simplifica como

$$\begin{aligned} A(z) &= \frac{1 - 8z + 27z^2 - 35z^3 + 14z^4}{(1-z)^2(1-2z)^2(1-3z)} \\ &= \frac{5/4}{1-z} + \frac{1/2}{(1-z)^2} - \frac{3}{1-2z} - \frac{2}{(1-2z)^2} + \frac{17/4}{1-3z} \end{aligned}$$

En consecuencia, tenemos

$$\begin{aligned} a_r &= \frac{5}{4} + \frac{1}{2}(r+1) - 3\cdot 2^r - 2(r+1)2^r + \frac{17}{4}3^r \\ &= \frac{7}{4} + \frac{r}{2} - r2^{r+1} - 5\cdot 2^r + \frac{17}{4}3^r \end{aligned}$$

Sugerimos al lector hacer un seguimiento de la derivación de la relación entre una función numérica y su función generadora.

Ejemplo 10.14

Consideremos una cierta reacción nuclear en el interior de un reactor que contiene núcleos y partículas nucleares de baja y alta energía. Existen dos tipos de eventos: 1) una partícula de alta energía choca con un núcleo y es absorbida, por lo que se origina la emisión de tres partículas de alta energía y una partícula de baja energía; 2) una partícula de baja energía choca con un núcleo y es absorbida, lo que origina la emisión de dos partículas de alta energía y una partícula de baja energía. Supongamos que cualquier partícula libre causará un evento 1 us después de ser emitida. También, que una sola partícula de alta energía se inyecta al tiempo 0 dentro de un sistema que contiene solamente núcleos. Queremos determinar el número de partículas de baja y alta energía en el sistema a un tiempo igual a r µs. Sean a_r y b_r el número de partículas de alta y baja energía en el tiempo r, respectivamente. Tenemos que

$$\begin{aligned} a_r &= 3a_{r-1} + 2b_{r-1} \\ b_r &= a_{r-1} + b_{r-1} \end{aligned} \tag{10.29}$$

Observemos que éstas son dos ecuaciones en diferencias simultáneas para las funciones numéricas a y b. Las ecuaciones son válidas para $r \geqq 1$, y tenemos las condiciones de frontera

$$a_0 = 1 \qquad b_0 = 0$$

A partir de (10.29) obtenemos

$$\sum_{r=1}^{\infty} a_r z^r = \sum_{r=1}^{\infty} 3a_{r-1} z^r + \sum_{r=1}^{\infty} 2b_{r-1} z^r$$

$$\sum_{r=1}^{\infty} b_r z^r = \sum_{r=1}^{\infty} a_{r-1} z^r + \sum_{r=1}^{\infty} b_{r-1} z^r$$

o

$$\begin{aligned} A(z) - 1 &= 3zA(z) + 2zB(z) \\ B(z) &= zA(z) + zB(z) \end{aligned} \tag{10.30}$$

De las ecuaciones (10.30) despejamos $A(z)$ y $B(z)$, y obtenemos

$$A(z) = \frac{1-z}{1-4z+z^2} = \frac{(3+\sqrt{3})/6}{1-(2+\sqrt{3})z} + \frac{(3-\sqrt{3})/6}{1-(2-\sqrt{3})z}$$

$$B(z) = \frac{z}{1-4z+z^2} = \frac{\sqrt{3}/6}{1-(2+\sqrt{3})z} - \frac{\sqrt{3}/6}{1-(2-\sqrt{3})z}$$

Y se tiene que

$$a_r = \frac{3+\sqrt{3}}{6}(2+\sqrt{3})^r + \frac{3-\sqrt{3}}{6}(2-\sqrt{3})^r$$

$$b_r = \frac{\sqrt{3}}{6}(2+\sqrt{3})^r - \frac{\sqrt{3}}{6}(2-\sqrt{3})^r$$

□

En el ejemplo 10.14 demostramos cómo podemos encontrar la solución de un conjunto de *ecuaciones en diferencias simultáneas* mediante la técnica de funciones generadoras. A partir de un conjunto de ecuaciones en diferencias simultáneas que relacionan varias funciones numéricas, podemos obtener un conjunto de ecuaciones (algebraicas) simultáneas que relacionen a sus funciones generadoras. Estas ecuaciones pueden resolverse para encontrar expresiones en forma cerrada para las funciones generadoras.

10.8 ALGORITMOS DE ORDENAMIENTO

Regresemos al ejemplo de la sección 10.2 en el cual mostramos que la función numérica $(3^0, 3^1, 3^2, \ldots, 3^r, \ldots)$ puede especificarse mediante la relación de recurrencia

$$a_r = 3a_{r-1} \tag{10.31}$$

con la condición de frontera $a_0 = 1$. La ecuación es una especificación *recursiva* del valor de a en *r*. Ésta dice que si a_{r-1} es conocido, entonces a_r puede calcularse como el triple de a_{r-1}. No obstante, si a_{r-2} es conocido, entonces a_{r-1} puede calcularse como el triple de a_{r-2}; si a_{r-3} es conocido, entonces a_{r-2} puede calcularse como el triple de a_{r-3} y así sucesivamente. Por tanto, conocer a_0 es suficiente para determinar a_r para cualquier *r*. Al recordar nuestro análisis de la sección 10.1, nos damos cuenta que tal punto de vista no está restringido a la descripción de funciones numéricas. Esto sugiere una manera concisa y estricta para describir algoritmos computacionales. En esta sección mostramos algunos ejemplos a partir del problema de ordenamiento. Como se estudió en la sección 8.2, el ordenamiento de *n* números que están almacenados en *n* registros $x_1, x_2, \ldots, x_n$, entendemos
el rearreglar el contenido de los registros de manera que los contenidos de los registros rearreglados $x_1, x_2, \ldots, x_n$ se encuentren en orden ascendente. Un método de ordenamiento consiste en una sucesión de pasos de comparación, cada uno de los cuales compara el contenido de los registros x_i y x_j, y coloca el número menor en x_i y el mayor en x_j. Denotaremos un paso de comparación por $A(x_i, x_j)$.

Primero presentaremos el algoritmo de ordenamiento de la sección 8.2, ORDENAMIENTOBURBUJA. Nuestra descripción del algoritmo será diferente a la presentada en la sección 8.2, en el sentido de que mostraremos una especificación recursiva del algoritmo.^ Introducimos una notación que nos permitirá dar una especificación "simbólica" del algoritmo. Sea $S(x_1, x_2, \ldots, x_n)$ el algoritmo de ORDENAMIENTOBURBUJA que ordena los números en los registros $x_1, x_2, \ldots, x_n$ en orden ascendente, y sea jel $M(x_1, x_2, \ldots, x_n)$ algoritmo MAYOR2 presentado en la sección 8.2 que coloca en el registro x_n al número mayor de entre los números de los registros $x_1, x_2, \ldots, x_n$. Tenemos que

$$S(x_1, x_2, \ldots, x_n) \triangleq M(x_1, x_2, \ldots, x_n)S(x_1, x_2, \ldots, x_{n-1}) \tag{10.32}$$

El símbolo 4 significa "está definido como" y la concatenación de los símbolos de los procedimientos significa la ejecución secuencial de los procedimientos correspondientes de izquierda a derecha. $M(x_1, x_2, \ldots, x_n)$ puede definirse como

$$M(x_1, x_2, \ldots, x_n) \triangleq A(x_1, x_2)A(x_2, x_3) \cdots A(x_{n-2}, x_{n-1})A(x_{n-1}, x_n)$$

lo cual es la especificación exacta del algoritmo MAYOR2 de la sección 8.2. La relación (10.32) puede escribirse como

$$\begin{aligned} S(x_1, x_2, \ldots, x_n) \triangleq A(x_1, x_2)A(x_2, x_3) \cdots \\ A(x_{n-2}, x_{n-1})A(x_{n-1}, x_n)S(x_1, x_2, \ldots, x_{n-1}) \end{aligned}$$

t Le recordamos al lector que el algoritmo de ordenamiento burbuja presentado aquí es el mismo que el presentado en la sección 8.2 en el sentido de que éste consiste en exactamente la misma sucesión de pasos de comparación. La única diferencia es la manera en que se especifican los pasos de comparación. En general, un algoritmo puede describirse de diferentes maneras, pero las más comunes son *no recursivos y recursivos.*

la cual es una descripción simbólica del algoritmo **ORDENAMIENTOBURBUJA.** Con la condición de frontera

$$S(x_1, x_2) \triangleq A(x_1, x_2)$$

nuestra especificación recursiva del algoritmo está completa. Como
un ejemplo explícito, observemos que

$$\begin{aligned} S(x_1, x_2, x_3, x_4) &\triangleq A(x_1, x_2)A(x_2, x_3)A(x_3, x_4)S(x_1, x_2, x_3) \\ &\triangleq A(x_1, x_2)A(x_2, x_3)A(x_3, x_4)A(x_1, x_2)A(x_2, x_3)S(x_1, x_2) \\ &\triangleq A(x_1, x_2)A(x_2, x_3)A(x_3, x_4)A(x_1, x_2)A(x_2, x_3)A(x_1, x_2) \end{aligned}$$

Sea a_n el número de pasos de comparación que el algoritmo de ordenamiento burbuja requiere para ordenar n números. De acuerdo con (10.32), tenemos que

$$a_n = (n-1) + a_{n-1} \tag{10.33}$$

Al hallar la solución de la relación de recurrencia en (10.33) y usando la condición de frontera $a_1 = 0$, obtenemos[1]

$$a_n = \frac{n(n-1)}{2}$$

Ahora presentaremos otro algoritmo de ordenamiento conocido como *algoritmo de Bose-Nelson,* hecho por R. C. Bose y R. J. Nelson (véase Bose y Nelson [2]), al cual denotaremos por $T(x_1, x_2, \ldots, x_n)$. Para simplificar nuestro análisis, supongamos que la cantidad de números a ser ordenados es una potencia de 2. Sea $P[(x_1, x_2, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{2n})]$ un algoritmo que arreglará los números en los registros $x_1, x_2, \ldots, x_{2n}$ en orden ascendente puesto que los números en los registros $x_1, x_2, \ldots, x_n$ ya habían sido arreglados en orden ascendente, al igual que los números en los registros $x_{n+1}, x_{n+2}, \ldots, x_{2n}$. Es claro,

$$\begin{aligned} T(x_1, x_2, \ldots, x_{2n}) \triangleq{}& T(x_1, x_2, \ldots, x_n)T(x_{n+1}, x_{n+2}, \ldots, x_{2n}) \\ & P[(x_1, x_2, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{2n})] \end{aligned} \tag{10.34}$$

En otras palabras, para ordenar los $2n$ números en los registros $x_1, x_2, \ldots, x_{2n}$, primero podemos ordenar los n números en $x_1, x_2, \ldots, x_n$, y los n números en $x_{n+1}, x_{n+2}, \ldots, x_{2n}$, y entonces combinar las dos sucesiones ordenadas en una. El algoritmo $P[(x_1, x_2, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{2n})]$ puede definirse recursivamente como:

$$\begin{aligned} & P[(x_1, x_2, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{2n})] \\ &\quad \triangleq P[(x_1, x_2, \ldots, x_{n/2}), (x_{n+1}, x_{n+2}, \ldots, x_{3n/2})] \\ &\qquad P[(x_{n/2+1}, x_{n/2+2}, \ldots, x_n), (x_{3n/2+1}, x_{3n/2+2}, \ldots, x_{2n})] \\ &\qquad P[(x_{n/2+1}, x_{n/2+2}, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{3n/2})] \end{aligned} \tag{10.35}$$

[1] Véase el problema 10.10.

Observemos que $P[(x_1, x_2, \ldots, x_{n/2}), (x_{n+1}, x_{n+2}, \ldots, x_{3n/2})]$ coloca a los *n/2* menores números, de entre los 2« números en los registros, en los registros $x_1, x_2, \ldots, x_{n/2}$ en orden ascendente. De modo similar, $P[(x_{n/2+1}, x_{n/2+2}, \ldots, x_{2n}), (x_{3n/2+1}, x_{3n/2+2}, \ldots, x_{2n})]$ coloca los *n/2* mayores números, de entre los 2« números en los registros, en los registros $x_{3n/2+1}, x_{3n/2+2}, \ldots, x_{2n}$ en orden ascendente. Por último. $P[(x_{n/2+1}, x_{n/2+2}, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{3n/2})$ I coloca los *n* restantes de los 2« números en los registros x_r $_{/2+1}, x_{n/2+2}, \ldots, x_{3n/2}$ en orden ascendente. Si aplicamos las relaciones (10.34) y (10.35) repetidamente, podemos expresar $T(x_1, x_2, \ldots, x_n)$, $n = 2^r$, como una sucesión de procedimientos de las formas T\x¡, $x_j)$ y $P[(x_i),(x_j)]$. No obstante, tanto $T(x_i, x_j)$ como $P[(x_i), (x_j)]$ son iguales a $A(x_i, x_j)$.

Por ejemplo, observemos que

$$T(x_1, x_2, x_3, x_4) \triangleq T(x_1, x_2)T(x_3, x_4)P[(x_1, x_2), (x_3, x_4)]$$

Puesto que

$$T(x_1, x_2) \triangleq A(x_1, x_2)$$
$$T(x_3, x_4) \triangleq A(x_3, x_4)$$

y

$$P[(x_1, x_2), (x_3, x_4)] \triangleq P[(x_1), (x_3)]P[(x_2), (x_4)]P[(x_2), (x_3)]$$
$$\triangleq A(x_1, x_3)A(x_2, x_4)A(x_2, x_3)$$

tenemos que

$$T(x_1, x_2, x_3, x_4) \triangleq A(x_1, x_2)A(x_3, x_4)A(x_1, x_3)A(x_2, x_4)A(x_2, x_3)$$

es una especificación completa de un algoritmo que ordena cuatro números almacenados en los registros x_1, x_2, x_3, x_4.

Sea c_r el número de pasos de comparación que realiza $P[(x_1, x_2, \ldots, x_{2^r}), (x_{2^r+1}, x_{2^r+2}, \ldots, x_{2^{r+1}}]$. De acuerdo con la relación en (10.35), tenemos que

$$c_r = 3c_{r-1} \tag{10.36}$$

Si hallamos la solución para (10.36) con la condición de frontera $c_0 = 1$, obtenemos

$$c_r = 3^r$$

Sea d_r el número de pasos de comparación que realiza el procedimiento $T(x_1, x_2, \ldots, x_{2^r})$. De acuerdo con la relación (10.34), tenemos que

$$d_r = 2d_{r-1} + c_{r-1}$$

o

$$d_r = 2d_{r-1} + 3^{r-1} \tag{10.37}$$

Si resolvemos (10.37), obtenemos

$$d_r = B2^r + 3^r$$

A partir de la condición de frontera $d_0 = 0$, determinamos que B = —\. Así,

$$d_r = 3^r - 2^r$$

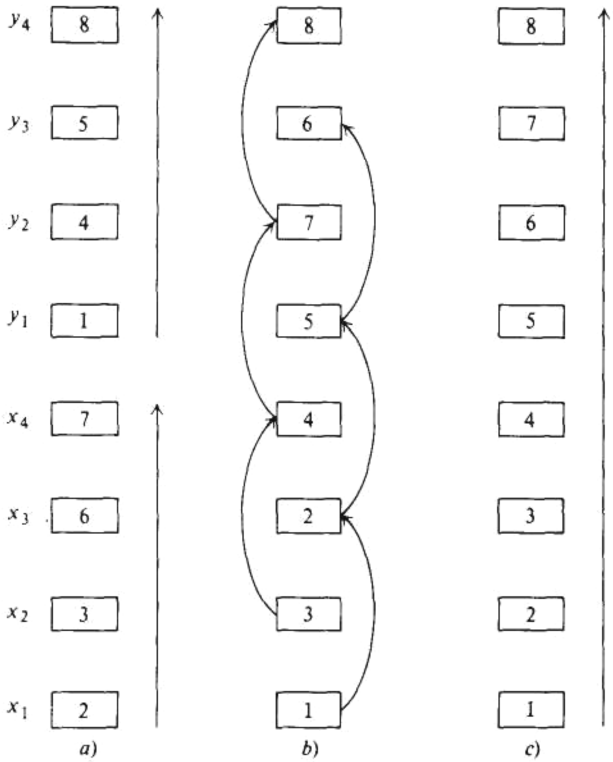


Figura 10.1

Otro algoritmo es el conocido como la *mezcla par-impar*, hecho por K.E. Batcher (véase Batcher [1]), al cual denotaremos por $U(x_1, x_2, \ldots, x_n)$. Suponemos que n es una potencia de 2. Sea

$$\begin{aligned} U(x_1, x_2, \ldots, x_{2n}) \triangleq\ & U(x_1, x_2, \ldots, x_n) U(x_{n+1}, x_{n+2}, \ldots, x_{2n}) \\ & B[(x_1, x_2, \ldots, x_n), (x_{n+1}, x_{n+2}, \ldots, x_{2n})] \end{aligned} \tag{10.38}$$

donde el algoritmo B se define como

$$\begin{aligned} & B[(x_1, x_2, x_3, x_4, \ldots, x_{2n}), (y_1, y_2, y_3, y_4, \ldots, y_{2n})] \\ & \quad \triangleq B[(x_1, x_3, x_5, \ldots, x_{2n-1}), (y_1, y_3, y_5, \ldots, y_{2n-1})] \\ & \quad\quad B[(x_2, x_4, x_6, \ldots, x_{2n}), (y_2, y_4, y_6, \ldots, y_{2n})] \\ & \quad\quad A(x_2, x_3) A(x_4, x_5) A(x_6, x_7) \cdots A(x_{2n}, y_1) \\ & \quad\quad A(y_2, y_3) A(y_4, y_5) \cdots A(y_{2n-2}, y_{2n-1}) \end{aligned} \tag{10.39}$$

No es obvio que el algoritmo definido en (10.39) mezcle dos secuencias ordenadas de números en una. Primero presentamos un ejemplo antes de proceder a verificar la validez del algoritmo. Supongamos que los ocho números, 2, 3, 6, 7, 1, 4, 5, 8, están almacenados en los registros $x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4$, como se muestra en la figura 10.1a, de tal manera que los números en los registros x_1, x_2, x_3, x_4 están en orden ascendente, y los números en los registros y_1, y_2, y_3, y_4, están en orden ascendente. De acuerdo con (10.39),

$$B[(x_1, x_2, x_3, x_4), (y_1, y_2, y_3, y_4)]$$
$$\triangleq B[(x_1, x_3), (y_1, y_3)]B[(x_2, x_4), (y_2, y_4)]A(x_2, x_3)A(x_4, y_1)A(y_2, y_3)$$

$B[(x_1, x_3), (y_1, y_3)]$ arregla los números en x_1, x_3, y_1, y_3 en orden ascendente, y $B[(x_2, x_4), (y_2, y_4)]$ arregla los números en x_2, x_4, y_2, y_4 en orden ascendente como se muestra en la figura 10.16. Finalmente, $A(x_2$, x3) arregla los números en x_2, x_3; $A(x_4, y_1)$ arregla los números en x_4, y_1; y $A(y_2, y_3)$ arregla los números en y_2, y_3 en orden ascendente como se muestra en la figura 10.1c.

Para verificar la validez de (10.39), primero queremos mostrar que después de la ejecución de $B[(x_1, x_3, x_5, \ldots, x_{2n-1}), (y_1, y_3, y_5, \ldots, y_{2n-1})]$, y $B[(x_2, x_4, x_6, \ldots, x_{2n}), (y_2, y_4, y_6, \ldots, y_{2n})]$ tenemos:

1. El número en x_1 es el menor de los *An* números.
2. El número en y_{2n} es el mayor de los *An* números.
3. Los dos números en x_{2i} y x_{2i+1}, $1 \leqq i \leqq n-1$, son el $2i$-ésimo y el $(2i+1)$-ésimo más pequeños de los *An* números; los dos números en x_{2n} y y_1 son el 2«-ésimo y el (2w + l)-ésimo más pequeños de los *An* números, y los dos números en y_{2i} y y_{2i+1}, $1 \leqq i \leqq$ *n*-\, son los (2n + 2/)-ésimo y *(2n* + 2/ + l)-ésimo más pequeños de los *An* números.

Si en efecto éste es el caso, las comparaciones. $A(x_2, x_3)A(x_4, x_5) \cdots A(y_{2n-2}, y_{2n-1})$ darán lugar a una sucesión ordenada de *An* números.

El lector puede aseverar que 1) y 2) son obvias. Para demostrar que 3) es verdadera, sólo mostraremos que dos números en x_{2i} y x_{2i+1}, $1 \leqq i \leqq n-1$, son el 2/-ésimo y el (2/ + l)-ésimo más pequeños de los *An* números; los otros dos casos pueden mostrarse en forma similar.[1] Consideremos el número que está en x_{2i} después de la ejecución. Este número es mayor que exactamente / - 1 de los números que estaban en $x_2, x_4, x_6, \ldots, x_{2n}, y_2, y_4, y_6, \ldots, y_{2n}$. Supongamos que este número es mayor que *k* de los números que estaban en $x_2, x_4, x_6, \ldots, x_{2n}$, y que es mayor que / - *k* - 1 de los números que estaban en $y_2, y_4, y_6, \ldots, y_{2n}$.
Esto es, este número es mayor que los números que estaban en $x_2, x_4, x_6, \ldots, x_{2k}$ y que los números que estaban en $y_2, y_4, y_6, \ldots, y_{2(i-k-1)}$. En otras palabras, este número es mayor que los números que estaban en $x_1, x_2, x_3, x_4, \ldots, x_{2k}$ y que los que estaban en $y_1, y_2, y_3, y_4, \ldots, y_{2(i-k-1)}$. Luego, si este número estaba en $x_{2(k+1)}$, también es mayor que x_{2k+1}, y podría ser o no ser mayor que el número que estaba en $y_{2(i-k-1)+1}$. Por otro lado, si este número estaba en $y_{2(i-k)}$, también es mayor que $y_{2(i-k)-1}$, y podría ser o no mayor que el número que estaba en x_{2k+1}. Por tanto, concluimos que este número es mayor que *2i* - 1 o 2/ de los *An* números en los registros.

De manera similar, consideremos el número que está en x_{2i+1} después de la ejecución. Suponemos que este número es mayor que *k* de los números que estaban en $x_1, x_3, \ldots,$

t En realidad no tenemos tres casos en (3). Sólo se debe a que en la notación que hemos escogido se tienen que hacer tres proposiciones paralelas.

x_{2n-1} y es mayor que $i - k$ de los números que estaban en $y_1, y_3, \ldots, y_{2n-1}$. Esto es, este número es mayor que los números que estaban en $x_1, x_3, \ldots, x_{2k-1}$, y es mayor que los números que estaban en $y_1, y_3, \ldots, y_{2(i-k)-1}$. Así, este número es mayor que los números que estaban en $x_1, x_2, x_3, \ldots, x_{2k-2}, x_{2k-1}$, y que los que estaban en $y_1, y_2, y_3, \ldots, y_{2(i-k)-2}, y_{2(i-k)-1}$. Luego, si este número estaba en. x_{2k+1}, también es mayor que el número que estaba en x_{2k}, y podría ser o no mayor que el número que estaba en $y_{2(i-k)}$. Si este número estaba en $y_{2(i-k)+1}$, también es mayor que el número que estaba en $y_{2(i-k)}$, y podría ser o no mayor que el número que estaba en x_{2k}. isn consecuencia, podemos concluir que este numero es mayor que $2i - 1$ o $2i$ de los 4« números en los registros.

Completamos la prueba de 3) al notar que si el número en x_{2i} es el $2i$-ésimo número más pequeño, entonces el número en x_{2i+1} debe ser *(2i* + l)-ésimo numero más pequeño, y viceversa.

Sea *c.* el número de pasos de comparación que *realiza* el algoritmo $B[(x_1, x_2, \ldots, x_{2^r}), (y_1, y_2, \ldots, y_{2^r})$. De acuerdo con la relación (10.39), tenemos que

$$c_r = 2c_{r-1} + (2^r - 1) \tag{10.40}$$

Al resolver (10.40), con la condición de frontera $c_0 = 1$, obtenemos que

$$c_r = r2^r + 1$$

Sea d_r el número de pasos de comparación que realiza el algoritmo $U(x_1, x_2, \ldots, x_{2^r})$. De acuerdo con la relación (10.38), tenemos que

$$d_r = 2d_{r-1} + c_{r-1}$$

o

$$d_r = 2d_{r-1} + (r-1)2^{r-1} + 1 \tag{10.41}$$

Al resolver la relación de recurrencia en (10.41), con la condición de frontera $d_Q = 0$, obtenemos

$$d_r = (r^2 - r + 4)2^{r-2} - 1$$

♦10.9 ALGORITMOS DE MULTIPLICACIÓN DE MATRICES

Como otra aplicación de algoritmos recursivos, consideremos el problema de multiplicar matrices. Sean A y B dos matrices de 2 x 2

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

El producto de A y B, denotado por C, puede calcularse como

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

donde

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} \\ c_{12} &= a_{11}b_{12} + a_{12}b_{22} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} \\ c_{22} &= a_{21}b_{12} + a_{22}b_{22} \end{aligned}$$

De inmediato observamos que el cálculo requiere cuatro operaciones de adición y ocho de multiplicación. Si d_x denota el costo de una operación de adición y d_2 el costo de una operación de multiplicación, el costo total de multiplicar dos matrices de 2 x 2 es

$$4d_1 + 8d_2$$

Una vez más, podemos preguntarnos si existe un algoritmo "mejor" para multiplicar dos matrices de 2 x 2. Resulta que un algoritmo alternativo puede hacerlo (más adelante, determinaremos si este algoritmo es efectivamente mejor). Dicho algoritmo calcula los primeros siete resultados intermedios:

$$\begin{aligned} m_1 &= (a_{11} + a_{22})(b_{11} + b_{22}) \\ m_2 &= (a_{12} - a_{22})(b_{21} + b_{22}) \\ m_3 &= (a_{11} - a_{21})(b_{11} + b_{12}) \\ m_4 &= (a_{11} + a_{12})b_{22} \qquad (10.42) \\ m_5 &= (a_{21} + a_{22})b_{11} \\ m_6 &= a_{11}(b_{12} - b_{22}) \\ m_7 &= a_{22}(b_{21} - b_{11}) \end{aligned}$$

Estos resultados intermedios pueden usarse para calcular las entradas de la matriz C:

$$\begin{aligned} c_{11} &= m_1 + m_2 - m_4 + m_7 \\ c_{12} &= m_4 + m_6 \\ c_{21} &= m_5 + m_7 \\ c_{22} &= m_1 - m_3 - m_5 + m_6 \end{aligned}$$

Un conteo rápido indica que ese algoritmo requiere 18 operaciones de adición y 7 de multiplicación. Así, el costo total es

$$18d_1 + 7d_2$$

Está muy lejos de ser evidente que ese algoritmo es superior al primer algoritmo, el cual es el algoritmo de multiplicación de matrices clásico. De hecho, uno podría señalar de inmediato que el primer algoritmo requiere 12 operaciones aritméticas, en tanto que el segundo requiere 25. Por otro lado, un partidario del segundo algoritmo podría argumentar que éste será menos costoso que el primero si una operación de multiplicación es significativamente más costosa que una operación de adición. Uno podría señalar que dicho argumento no es real debido a que, en la práctica, el costo de la multiplicación y la adición es semejante. No obstante, antes de que concluyamos que el diseño de este algoritmo alternativo es un ejercicio vacío, consideremos el problema más general de multiplicar dos matrices A y B de *n* x *n*. Para simplificar nuestro análisis, supondremos que *n* es una potencia de 2. Recordemos que el cálculo del producto de dos matrices de *n* x *n* mediante el algoritmo de multiplicación clásico requiere «^{3} operaciones de multiplicación y *n\n* - 1) operaciones de adición, esto es, un total de 2«^{3} - n^2 operaciones aritméticas (la y-ésima entrada del producto se calcula como $\Sigma_{k=1}^{n} a_{ik}b_{kj}$, lo cual requiere *n* operaciones de multiplicación y *(n* - 1) operaciones de adición). Nos damos cuenta que el segundo algoritmo presentado puede usarse para multiplicar dos matrices de *n* x *n.* Dadas dos matrices de *n* x « A y B, podemos particionarlas en las submatrices de *(n/2)* x («/2):

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \hline \mathbf{a}_{21} & \mathbf{a}_{22} \end{array}\right]$$

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{b}_{11} & \mathbf{b}_{12} \\ \hline \mathbf{b}_{21} & \mathbf{b}_{22} \end{array}\right]$$

Entonces, el producto AB puede expresarse como

$$\mathbf{C} = \mathbf{AB} = \left[\begin{array}{c|c} \mathbf{c}_{11} & \mathbf{c}_{12} \\ \hline \mathbf{c}_{21} & \mathbf{c}_{22} \end{array}\right]$$

donde $\mathbf{a}_{11}$, $\mathbf{a}_{12}$, $\mathbf{a}_{21}$, $\mathbf{a}_{22}$, $\mathbf{b}_{11}$, $\mathbf{b}_{12}$, $\mathbf{b}_{21}$, $\mathbf{b}_{22}$, $\mathbf{c}_{11}$, $\mathbf{c}_{12}$, $\mathbf{c}_{21}$ y $\mathbf{c}_{22}$[†] son todas submatrices de $(n/2) \times (n/2)$ Entonces podemos calcular

$$\begin{aligned}
\mathbf{m}_1 &= (\mathbf{a}_{11} + \mathbf{a}_{22})(\mathbf{b}_{11} + \mathbf{b}_{22}) \\
\mathbf{m}_2 &= (\mathbf{a}_{12} - \mathbf{a}_{22})(\mathbf{b}_{21} + \mathbf{b}_{22}) \\
\mathbf{m}_3 &= (\mathbf{a}_{11} - \mathbf{a}_{21})(\mathbf{b}_{11} + \mathbf{b}_{12}) \\
\mathbf{m}_4 &= (\mathbf{a}_{11} + \mathbf{a}_{12})\mathbf{b}_{22} \\
\mathbf{m}_5 &= (\mathbf{a}_{21} + \mathbf{a}_{22})\mathbf{b}_{11} \\
\mathbf{m}_6 &= \mathbf{a}_{11}(\mathbf{b}_{12} - \mathbf{b}_{22}) \\
\mathbf{m}_7 &= \mathbf{a}_{22}(\mathbf{b}_{21} - \mathbf{b}_{11})
\end{aligned}$$

t Usamos el tipo en negritas para indicar que son matrices.

Observemos que éstas son repeticiones de las relaciones en (10.42). Únicamente hemos remplazado las letras normales por negritas para enfatizar que $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_7, \mathbf{a}_{11}, \ldots, \mathbf{a}_{22}, \ldots, \mathbf{b}_{11}, \ldots, \mathbf{b}_{22}$ son todas matrices. Desde luego, en estas ecuaciones todas las multiplicaciones lo son de matrices de *(n/2)* x *(n/2),* y todas las adiciones lo son de matrices de *(n/2)* x *(n/2).* Finalmente, podemos calcular

$$
\begin{aligned}
\mathbf{c}_{11} &= \mathbf{m}_1 + \mathbf{m}_2 - \mathbf{m}_4 + \mathbf{m}_7 \\
\mathbf{c}_{12} &= \mathbf{m}_4 + \mathbf{m}_6 \\
\mathbf{c}_{21} &= \mathbf{m}_5 + \mathbf{m}_7 \\
\mathbf{c}_{22} &= \mathbf{m}_1 - \mathbf{m}_3 - \mathbf{m}_5 + \mathbf{m}_6
\end{aligned}
$$

Observemos que el costo total de multiplicar A y B es de 7 multiplicaciones matriciales y 18 adiciones matriciales (de matrices de *(n/2)* x *(n/2)).* Si utilizamos el método clásico para llevar a cabo la multiplicación de las matrices de *(n/2)* x *(n/2),* el número total de operaciones aritméticas será

$$7 \times \left[2\left(\frac{n}{2}\right)^3 - \left(\frac{n}{2}\right)^2 \right] + 18 \times \left(\frac{n}{2}\right)^2 = \frac{7}{4} n^3 + \frac{11}{4} n^2$$

que es superior a 2«^{3} - n^2 para *n* grande. No obstante, podemos mejorar lo anterior mediante el empleo recursivo de nuestro nuevo algoritmo de multiplicación para calcular los productos de las submatrices. En otras palabras, para calcular el producto de dos matrices de *n* x «, necesitamos realizar 7 multiplicaciones y 18 adiciones de matrices de *(n/2)* x *(n/2),* de las cuales, cada una de las operaciones de multiplicación de matrices de *(n/2)* x *(n/2),* puede realizarse con 7 multiplicaciones y 18 adiciones de matrices de *(niA)* x *(niA),* y así sucesivamente. Sea/$_r$ el número total de operaciones aritméticas requeridas para multiplicar dos matrices de 2^r x 2^r, tenemos que:

$$f_r = 7f_{r-1} + 18 \cdot 2^{2r-2} \qquad r \geqq 1 \qquad (10.43)$$

con la condición de frontera/$_0$ = 1. Al resolver la ecuación de (10.43), obtenemos que:

$$f_r = 7 \cdot 7^r - \tfrac{18}{3} \cdot 4^r$$

Así, para *n* - 2^r,

$$f_r = 7 \cdot 7^{\lg n} - \tfrac{18}{3} \cdot n^2 = 7n^{2.81} - \tfrac{18}{3} \cdot n^2$$

Puesto que la complejidad temporal del algoritmo presentado arriba es ®(«^{281}), la complejidad temporal del problema de multiplicación de matrices es *O(n*$^{2\ 81}$).

10.10 NOTAS Y REFERENCIAS

Véase Levy y Lessman [4] en lo referente a ecuaciones en diferencias finitas. También véanse los capítulos 2 y 3 de Liu y Liu [6]. Los algoritmos de ordenamiento analizados en la sección 10.8 tienen dos características distintas: 1) son *no-adaptivos* en el sentido de que la sucesión de pasos de comparación en un algoritmo está predeterminada y no varía en concordancia con el resultado de ningún paso de comparación particular, y 2) no se necesitan registros adicionales para almacenar los resultados intermedios. Observemos que en un algoritmo no-adaptivo, algunos de los pasos de comparación pueden llevarse a cabo simultáneamente [por ejemplo, $A(x_1, x_2)$ y $A(x_3, x_4)$ pueden llevarse de modo simultáneo en tanto que $A(x_1, x_2)$ y $A(x_1, x_3)$ no]. En consecuencia, existe la posibilidad de incrementar la rapidez del cálculo mediante procesamiento en paralelo. La ventaja de algoritmos que no usan registros de almacenamiento adicionales es obvia cuando el número de objetos a ser almacenados es grande. Para un análisis más completo sobre algoritmos de ordenamiento y tópicos relacionados, véase el capítulo 5 de Knuth [3].

El algoritmo de multiplicación matricial presentado en la sección 10.9 fue tomado de Strassen [8]. Véase el problema 10.38 para un algoritmo que emplea 7 operaciones de multiplicación y 15 operaciones de adición. Observemos, no obstante, que la complejidad temporal de dicho algoritmo podría continuar siendo $\Theta(n^{2.81})$. Para reducir el exponente de n desde 2.81 a un número menor, la reacción inmediata de uno podría ser la búsqueda de un algoritmo de multiplicación de matrices de 3 x 3 que utilice 21 o menos operaciones de multiplicación. Sin embargo, no ha sido descubierto hasta el momento. Resulta sorprendente que los algoritmos de multiplicación para matrices muy grandes que conducen a un mejoramiento del exponente de n ya han sido descubiertos. Por ejemplo, Pan [7] muestra que podemos multiplicar dos matrices de 48 x 48 por medio de 47 126 operaciones de multiplicación, lo cual reduce el exponente a 2.78.

1. Batcher, K. E.: Sorting Networks and Their Applications, *AFIPS Proc. ofthe ¡968 SJCS,* 32: 307-314(1968).
2. Bose, R. C, y R. J. Nelson: "A Sorting Problem", *Journal ofACM,* 9: 282-296 (1962).
3. Knuth, D. E.: *The Art ofComputer Programming, Vol. 3, Sorting and Searching,* Addison-Wesley Publishing Company, Reading, Mass., 1973.
4. Levy, H. y F. Lessman: *Finite Difference Equations,* The Macmillan Company, Nueva York, 1961.
5. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
6. Liu, C. L. y J. W. S. Liu: *Linear Systems Analysis,* McGraw-Hill Book Company, Nueva York, 1975.
7. Pan, V. Ya.: "Strassen's Algorithm Is Not Optimal", *Proc. 19th AnnualSymposium on Foundations of Computer Science,* 166-176 (1978).
8. Strassen, V: "Gaussian Elimination Is Not Optimal", *Numerische Mathematik,* 13: 354-356 (1969).

PROBLEMAS

10.1 Encuentre la solución de las siguientes relaciones de recurrencia:

a) $a_r - 7a_{r-1} + 10a_{r-2} = 0$, dado que $a_0 = 0$ y $a_1 = 3$.

b) $a_r - 4a_{r-1} + 4a_{r-2} = 0$, dado que $a_0 = 1$ y $a_1 = 6$.

10.2 Encuentre la solución de las siguientes relaciones de recurrencia:

a) $a_r - 7a_{r-1} + 10a_{r-2} = 3^r$, dado que $a_0 = 0$ y $a_1 = 1$.

b) $a_r + 6a_{r-1} + 9a_{r-2} = 3$, dado que $a_0 = 0$ y $a_1 = 1$.

c) $a_r + a_{r-1} + a_{r-2} = 0$, dado que $a_0 = 0$ y $a_1 = 2$.

10.3 Encuentre la solución de las siguientes relaciones de recurrencia:

a) $a_r - a_{r-1} - a_{r-2} = 0$, dado que $a_0 = 1$ y $a_1 = 1$.

b) $a_r - 2a_{r-1} + 2a_{r-2} - a_{r-3} = 0$, dado que $a_0 = 2$, $a_1 = 1$ y $a_2 = 1$.

10.4 Dado que $a_0 = 0$, $a_1 = 1$, $a_2 = 4$ y $a_3 = 12$ satisfacen la relación de recurrencia

$$a_r + C_1 a_{r-1} + C_2 a_{r-2} = 0$$

determine a_r.

10.5 La solución de la relación de recurrencia

$$C_0 a_r + C_1 a_{r-1} + C_2 a_{r-2} = f(r)$$

es

$$3^r + 4^r \div 2$$

Dado que $f(r) = 6$ para todo r, determine C_0, C_1 y C_2.

10.6 La solución de la relación de recurrencia

$$a_r = A a_{r-1} + B 3^r \qquad r \geqq 1$$

es

$$a_r = C 2^r + D 3^{r+1} \qquad r \geqq 0$$

Dados $a_0 = 19$ y $a_1 = 50$, determine las constantes *A, B, C* y *D.*

10.7 Sea

$$4a_r + C_1 a_{r-1} + C_2 a_{r-2} = f(r) \qquad r \geqq 2$$

una relación de recurrencia lineal de segundo orden con coeficientes constantes. Para algunas condiciones de frontera a_Q y a_u la solución de la relación de recurrencia es

$$1 - 2r + 3 \cdot 2^r$$

Determine a_0, a_1, C_1, C_2 y $f(r)$ (la solución no es única).

10.8 Considere la relación de recurrencia

$$a_r = a_{r-1} - a_{r-2}$$

a) Encuentre la solución de la relación de recurrencia, ya que $a_x = 1$ y $a_2 = 0$.
b) ¿Puede encontrar la solución de la relación de recurrencia si se tiene que $a_0 = 0$ y $<z_3 = 0$?
c) Repita el inciso *b)* si se tiene que $a_0 = 1$ y $a_3 = 2$.

10.9 *a)* Determine la solución particular para la ecuación en diferencias

$$a_r - 3a_{r-1} + 2a_{r-2} = 2^r$$

b) Determine la solución particular para la ecuación en diferencias

$$a_r - 4a_{r-1} + 4a_{r-2} = 2^r$$

10.10 *a)* Determine la solución particular para la ecuación en diferencias

$$a_r - 2a_{r-1} = f(r)$$

donde $f(r) = 7r$.
b) Repita el inciso *a)* para. $f(r) = 7r^2$.
c) Determine la solución particular para la ecuación en diferencias

$$a_r - a_{r-1} = 7r$$

d) Repita el inciso *c)* si $f(r) = 7r^2$.
e) Sea

$$C_0 a_r + C_1 a_{r-1} + \cdots + C_k a_{r-k} = f(r)$$

una ecuación en diferencias con una raíz característica 1. Sea $f(r) = r^4$. ¿Qué puede decir acerca de la forma general de la solución particular $a_r^{(p)}$?

10.11 *a)* Halle la solución de la relación de recurrencia

$$a_r + 3a_{r-1} + 2a_{r-2} = f(r)$$

donde

$$f(r) = \begin{cases} 1 & r = 2 \\ 0 & \text{de otra manera} \end{cases}$$

con la condición de frontera $a_0 = a_1 = 0$.

b) Repita el inciso *a)* para

$$f(r) = \begin{cases} 1 & r = 5 \\ 0 & \text{de otra manera} \end{cases}$$

c) Considere la relación de recurrencia

$$C_0a_r + C_1a_{r-1} + C_2a_{r-2} + \cdots + C_ka_{r-k} = f(r)$$

Sea $\hat{a}_r$ la solución de la relación de recurrencia para $f(r)=\hat{f}(r)$ con las condiciones de frontera $\hat{a}_0 = \hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_{k-1} = 0$. Sea $\tilde{a}_r$ la solución de la relación de recurrencia para $f(r) = \tilde{f}(r)$ con las condiciones de frontera $\tilde{a}_0 = \tilde{a}_1 = \tilde{a}_2 = \cdots = \tilde{a}_{k-1} = 0$. Dado $\hat{f}(r) = 0$ para $r < k$, y

$$\tilde{f}(r) = \begin{cases} 0 & 0 \leqq r \leqq l-1 \\ \hat{f}(r-l) & r \geqq l \end{cases}$$

para algún / fijo, ¿qué se puede concluir acerca de $\hat{a}_r$ y $\tilde{a}_r$?

10.12 *a)* Considere la relación de recurrencia

$$C_0a_r + C_1a_{r-1} + C_2a_{r-2} + \cdots + C_ka_{r-k} = f(r)$$

Sea $\hat{a}_r$ la solución de la relación de recurrencia para $f(r)=\hat{f}(r)$ con la condiciones de frontera $\hat{a}_0 = \hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_{k-1} = 0$. Sea $\tilde{a}_r$ la solución de la relación de recurrencia para $f(r) = \tilde{f}(r)$ con las condiciones de frontera $\tilde{a}_0 = \tilde{a}_1 = \tilde{a}_2 = \cdots = \tilde{a}_{k-1} = 0$. Demuestre que $\bar{a}_r = \hat{a}_r + \tilde{a}_r$ es la solución de la relación de recurrencia para $f(r) = \hat{f}(r) + \tilde{f}(r)$ con las condiciones de frontera $\bar{a}_0 = \bar{a}_1 = \bar{a}_2 = \cdots = \bar{a}_{k-1} = 0$, siempre que $\hat{f}(r) = \tilde{f}(r) = 0$ para $r < k$.

b) Encuentre la solución de la relación de recurrencia

$$a_r + 5a_{r-1} + 6a_{r-2} = f(r)$$

donde

$$f(r) = \begin{cases} 0 & r = 0,\ 1,\ 5 \\ 6 & \text{de otra manera} \end{cases}$$

yaque $a_0 = a_1 = 0$.

10.13 Un rumor se difunde vía telefónica entre *r* personas. En específico en una conversación poi teléfono entre *A y B,Ale* cuenta a *B* todo el rumor que ha escuchado, y *B* también le cuenta a *A* Sea a_r el número mínimo de llamadas telefónicas entre *r* personas de manera que todo el rumoi será conocido por todas las personas.

a) Demuestre que, $a_2 = 1$, $a_3 = 3$ y $a_4 = 4$.

b) Demuestre que

$$a_r \leqq a_{r-1} + 2$$

c) Demuestre que

$$a_r \leqq 2r - 4 \qquad \text{para } r \geqq 4$$

[En efecto, puede demostrarse que $a_r = 2r - 4$. Véase, B. Baker y R. Shostak, Rumores y teléfonos. *Discrete Mathematics*, **2**: 191-193, (1972).]

10.11 *a)* Halle la solución de la relación de recurrencia

$$a_r + 3a_{r-1} + 2a_{r-2} = f(r)$$

donde

$$f(r) = \begin{cases} 1 & r = 2 \\ 0 & \text{de otra manera} \end{cases}$$

con la condición de frontera $a_0 = a_1 = 0$.

b) Repita el inciso *a)* para

$$f(r) = \begin{cases} 1 & r = 5 \\ 0 & \text{de otra manera} \end{cases}$$

c) Considere la relación de recurrencia

$$C_0a_r + C_1a_{r-1} + C_2a_{r-2} + \cdots + C_ka_{r-k} = f(r)$$

Sea $\hat{a}_r$ la solución de la relación de recurrencia para $f(r) = \hat{f}(r)$ con las condiciones de frontera $\hat{a}_0 = \hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_{k-1} = 0$. Sea $\tilde{a}_r$ la solución de la relación de recurrencia para $f(r) = \tilde{f}(r)$ con las condiciones de frontera $\tilde{a}_0 = \tilde{a}_1 = \tilde{a}_2 = \cdots = \tilde{a}_{k-1} = 0$. Dado $\hat{f}(r) = 0$ para $r < k$, y

$$\tilde{f}(r) = \begin{cases} 0 & 0 \leq r \leq l-1 \\ \hat{f}(r-l) & r \geq l \end{cases}$$

para algún / fijo, ¿qué se puede concluir acerca de $\hat{a}_r$ y $\tilde{a}_r$?

10.12 *a)* Considere la relación de recurrencia

$$C_0a_r + C_1a_{r-1} + C_2a_{r-2} + \cdots + C_ka_{r-k} = f(r)$$

Sea $\hat{a}_r$ la solución de la relación de recurrencia para $f(r) = \hat{f}(r)$ con la condiciones de frontera $\hat{a}_0 = \hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_{k-1} = 0$. Sea $\tilde{a}_r$ la solución de la relación de recurrencia paraj $f(r) = \tilde{f}(r)$ con las condiciones de frontera $\tilde{a}_0 = \tilde{a}_1 = \tilde{a}_2 = \cdots = \tilde{a}_{k-1} = 0$. Demuestre que $a_r = \hat{a}_r + \tilde{a}_r$ es la solución de la relación de recurrencia para; $f(r) = \hat{f}(r) + \tilde{f}(r)$ con las condiciones de frontera $\bar{a}_0 = \bar{a}_1 = \bar{a}_2 = \cdots = \bar{a}_{k-1} = 0$, siempre que $\hat{f}(r) = \tilde{f}(r) = 0$ para $r < k$.

b) Encuentre la solución de la relación de recurrencia

$$a_r + 5a_{r-1} + 6a_{r-2} = f(r)$$

donde

$$f(r) = \begin{cases} 0 & r = 0,\ 1,\ 5 \\ 6 & \text{de otra manera} \end{cases}$$

yaque $a_0 = a_1 = 0$.

10.13 Un rumor se difunde vía telefónica entre *r* personas. En específico en una conversación poi teléfono entre *A y B,Ale* cuenta a *B* todo el rumor que ha escuchado, y *B* también le cuenta a *A* Sea a_r el número mínimo de llamadas telefónicas entre *r* personas de manera que todo el rumoi será conocido por todas las personas.

a) Demuestre que, $a_2 = 1$, $a_3 = 3$ y $a_4 = 4$.

b) Demuestre que

$$a_r \leq a_{r-1} + 2$$

c) Demuestre que

$$a_r \leq 2r - 4 \qquad \text{para } r \geq 4$$

[En efecto, puede demostrarse que $a_r = 2r - 4$. Véase, B. Baker y R. Shostak, Rumores y teléfonos. *Discrete Mathematics,* **2**: 191-193, (1972).]

10.14 Si a_r denota el número de particiones de un conjunto de r elementos, demuestre que

$$a_{r+1} = \sum_{i=0}^{r} \binom{r}{i} a_i$$

donde $a_0 = 1$.

10.15 Considere un sistema de control de tráfico aéreo en el cual la altitud deseada de una aeronave, a_r, se calcula cada segundo con una computadora y se compara con la altitud real de la aeronave, b_{r-1}, es determinada por un radar de localización un segundo antes. Dependiendo si a_r es mayor o menor que b_{r-1}, la altitud de la aeronave se cambiará en la forma correspondiente. El cambio en la altitud en el r-ésimo segundo, $b_r - b_{r-1}$; es proporcional a la diferencia $a_r - b_{r-1}$. Esto es,

$$b_r - b_{r-1} = K(a_r - b_{r-1})$$

donde K es la constante de proporcionalidad.

a) Determine b_r, ya que $a_r = 1\,000\left(\frac{3}{2}\right)^2$, $K = 3$, y $b_0 = 0$.

b) Determine b_r, ya que

$$a_r = \begin{cases} 1000\left(\frac{3}{2}\right)^r & 0 \leqq r \leqq 9 \\ 1000\left(\frac{3}{2}\right)^{10} & r \geqq 10 \end{cases}$$

$K = 3$ y $b_0 = 0$.

10.16 *El problema de la torre de Hanoi,* r discos de tamaños decrecientes y con un orificio en su centro se deslizan sobre un poste con el disco mayor en el fondo, como se muestra en la figura 10P.1. Estos discos se transfieren uno por uno hacia otro poste, y existe un tercer poste disponible sobre el cual pueden colocarse los discos temporalmente. Si, durante la transferencia de los discos, ninguno puede colocarse sobre otro disco menor, ¿en cuántos movimientos podrán ser transferidos los discos sin cambiar sus posiciones relativas?

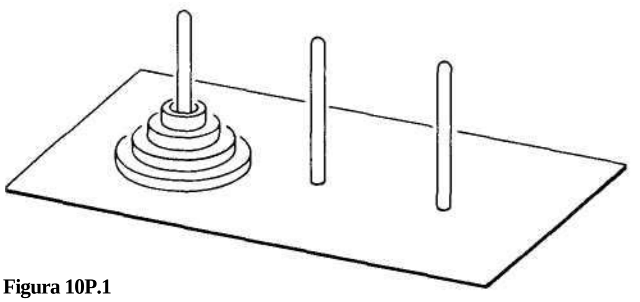


Figura 10P.1

10.17 Considere la multiplicación de bacterias en un ambiente controlado. El número de bacterias que hay en el r-ésimo día es denotado por a_r. Defina la razón de crecimiento en el r-ésimo día como $a_r - 2a_{r-1}$. Si sabe que la razón de crecimiento se duplica cada día, determine a_r, ya que $a_0 = 1$.

10.18 Considere la operación de una fábrica cuyo promedio de utilidades cada dos meses sucesivos es igual al promedio de las nuevas órdenes en ese periodo. Sean a, la nueva orden recibida y b_r la utilidad mensual en el r-ésimo mes.

a) Dado que $a_r = 2^r$ para toda $r \geqq 0$ y $b_0 = 0$, determine b_r.

b) Repita el inciso *a)* para

$$a_r = \begin{cases} 2^r & 0 \leqq r \leqq 9 \\ 2^{10} & r \geqq 10 \end{cases}$$

10.19 Una partícula se mueve en dirección horizontal. La distancia que recorre a cada segundo es igual al doble de la distancia que recorrió en el segundo anterior. Sea a_r la posición de la partícula en el r-ésimo segundo. Determine a_r, ya que $a_0 = 3$ y $a_3 = 10$.

10.20 Considere el siguiente algoritmo para ordenar r números para $r \geqq 2$.

d) Utilice $2r - 3$ comparaciones para determinar el número mayor y el menor inmediato a éste de los r números. *b)* Ordene recursivamente los r - 2 números restantes.

Sea a_r el número de comparaciones usadas para ordenar r números. Determine a_r

10.21 Sea a_r el número de aristas de un grafo completo de r vértices.

a) Obtenga una relación de recurrencia para a_r en términos de a_{r-1}.

b) Halle la solución de la relación de recurrencia.

10.22 Sea a_r el número de subconjuntos del conjunto $\{1, 2, \ldots, r\}$ que no contienen dos números consecutivos. Determine a_r.

Sugerencia: entre los a_r subconjuntos, ¿cuántos de ellos no contienen al número r?, ¿cuántos de ellos contienen al número *rl*

10.23 Considere el problema de cubrir una tira rectangular de longitud *n* con dos tipos de fichas de dominó. Una ficha azul tiene longitud 2, mientras que la ficha naranja tiene longitud 1. Conside rando que hay suficientes fichas de cada tipo, ¿de cuántas maneras puede ser cubierta la tira?

10.24 ¿Cuántas sucesiones binarias de r-dígitos que no tengan números 0 adyacentes existen?

10.25 Sea a_r el activo total de un banco al final del r-ésimo mes, el cual es igual a la suma del depósito total en el r-ésimo mes y 1.1 veces el total de activo al final del *(r* - l)-ésimo mes. Debido a que el total del depósito es una constante 100 (en miles de dólares), determine a_r si $a_0 = 0$.

10.26 Sea a_r el total de dólares activos de una compañía en el r-ésimo año. Es obvio que $a_r - a_{r-1}$ es el incremento en el activo durante el r-ésimo año. Si el incremento en activos durante cada año siempre es cinco veces el incremento del año anterior, ¿cuáles son los activos totales en el r-ésimo año? Se sabe que $a_0 = 3$ y $a_1 = 7$.

10.27 Sea a_r el número de regiones sin traslape en las cuales es dividido el interior de un polígono convexo de r lados por sus diagonales. Suponga que no es posible que tres diagonales se corten en un punto.

a) Demuestre que

$$a_r - a_{r-1} = \frac{(r-1)(r-2)(r-3)}{6} + r - 2 \qquad r \geqq 3$$

y $a_0 = a_1 = a_2 = 0$.

b) Determine *A(z).*

c) Determine a_r a partir de *A(z).*

10.28 Existen dos tipos de partículas dentro de un reactor nuclear. A cada segundo una partícula a se dividirá en tres partículas β, y una partícula β se dividirá en una partícula a y dos partículas β. Si sólo hay una partícula a en el reactor en / = 0, ¿cuántas partículas hay en total para t = 100?

10.29 ¿Cuántos árboles generadores tiene el grafo escalera de la figura 10P.2?

Sugerencia: si a_r denota el número de árboles generadores que tiene el grafo escalera de r escalones, y si b_r denota el número de árboles generadores que incluyen el primer escalón que tiene el grafo escalera de r escalones, exprese a_r en términos de a_{r-1} y b_r. Exprese b_r en términos de a_{r-1} y b_{r-1}.

Figura 10P.2

10.30 Sea d_r el número de maneras de permutar r enteros $\{1, 2, \ldots, r\}$ ....de manera que el entero i no estará en la ¿-ésima posición para $1 \leqq i \leqq r$.

a) Utilice un argumento combinatorio para demostrar que

$$d_r = (r-1)(d_{r-1} + d_{r-2})$$

b) Demuestre que

$$d_r = r!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^r \frac{1}{r!}\right]$$

satisface la relación de recurrencia del inciso *a).*

10.31. Una sucesión de dígitos binarios es alimentada en un contador a una razón de 1 digito/s. El contador está diseñado para registrar 1 s en la sucesión de entrada. Sin embargo, éste es tan lento que se cierra exactamente 7 s después de cada registro, tiempo durante el cual los dígitos de entrada son ignorados. Sea a_r el número de sucesiones binarias de longitud r al final de las cuales el contador no se encuentra cerrado. Hallar la ecuación en diferencias a la cual satisface a_n las condiciones de frontera, y la función generadora *A(z).*

10.32 Para el grafo de la figura 10P.3, determine el número de paseos dirigidos de longitud n que inician a partir del vértice a y terminan en el vértice d.

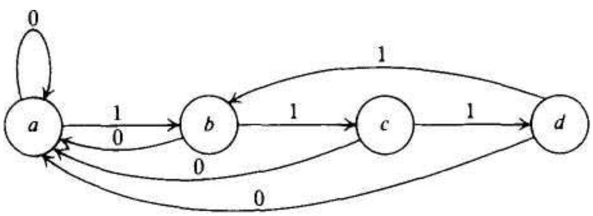


Figura 10P.3

10.33 *a)* Sea *A(z)* la función generadora de una función numérica a. Demuestre que

$$a_0 = \lim_{z \to 0} A(z)$$

(Este resultado se conoce como el *teorema del valor inicial?) b)* Verifique el teorema del valor inicial para la función numérica a, donde $a_r = 2^{r+1}$ para toda r.

10.34 Encuentre la solución de la ecuación en diferencias

$$a_r^2 - 2a_{r-1}^2 = 1$$

dado que $a_0 = 2$.

Sugerencia: considere que $b_r = a_r^2$. **10.35** Halle la solución de la ecuación en diferencias

$$ra_r + ra_{r-1} - a_{r-1} = 2^r$$

dado que $a_0 = 273$.

Sugerencia: considere $b_r = ra_r$. **10.36** *a)* Encuentre la solución de la ecuación en diferencias

$$a_r^2 - 2a_{r-1} = 0$$

dado que $a_0 = 4$.

Sugerencia: considere que $b_r = \lg a_r$, donde lg denota el logaritmo base 2.

b) Halle la solución de la ecuación en diferencias
dado que $a_0 = 4$.

Sugerencia:

$$a_r = \sqrt{a_{r-1} + \sqrt{a_{r-2} + \sqrt{a_{r-3} + \sqrt{\cdots}}}}$$

calcule *a}*.

10.37 Encuentre la solución de la ecuación en diferencias

$$a_r - ra_{r-1} = r! \qquad \text{para } r \geqq 1$$

dado que $a_0 = 2$.
Sugerencia: considere que b_r = *ajr*.

10.38 El producto de dos matrices A y B de 2 x 2 se puede calcular usando 7 multiplicaciones y 15 adiciones. Primero se calculan los siguientes resultados intermedios:

$s_1 = a_{21} + a_{22}$	$m_1 = s_2 s_6$	$t_1 = m_1 + m_2$
$s_2 = s_1 - a_{11}$	$m_2 = a_{11} b_{11}$	$t_2 = t_1 + m_4$
$s_3 = a_{11} - a_{21}$	$m_3 = a_{12} b_{21}$	
$s_4 = a_{12} - s_2$	$m_4 = s_3 s_7$	
$s_5 = b_{12} - b_{11}$	$m_5 = s_1 s_5$	
$s_6 = b_{22} - s_5$	$m_6 = s_4 b_{22}$	
$s_7 = b_{22} - b_{12}$	$m_7 = a_{22} s_8$	
$s_8 = s_6 - b_{21}$		

Los elementos de la matriz producto **C** pueden calcularse como:

$$c_{11} = m_2 + m_3$$
$$c_{12} = t_1 + m_5 + m_6$$
$$c_{21} = t_2 - m_7$$
$$c_{22} = t_2 + m_5$$

Confirme que estas ecuaciones dan lugar a los resultados correctos.

CAPÍTULO

ONCE

Grupos y anillos

11.1 INTRODUCCIÓN

Consideremos el funcionamiento de una máquina vendedora[1'] que entrega un paquete de goma de mascar cuando se depositan dos monedas de diez centavos, una barra de caramelo cuando se depositan dos monedas de diez centavos y una de veinticinco centavos, y un paquete de cigarros cuando se depositan dos monedas de veinticinco centavos. Sean A = {moneda de 10¢, moneda de 25¢} y B = {goma de mascar, caramelo, cigarros}. El funcionamiento de la máquina vendedora puede describirse formalmente como una función de $A \times A$ hacia B, la cual se muestra en la figura 11.1a. La tabla de la figura 11.1a puede establecerse de manera alternativa como se muestra en la figura 11.1 ¿>, donde las filas indican la primera moneda depositada y las columnas indican la segunda.

Otro ejemplo, supongamos que el color del cabello de un niflo está determinado por el cabello de sus padres, como se muestra en la figura 11.2. La relación entre el color del cabello de un niño y el de sus padres puede describirse mediante una función *de AxA* hacia.4, donde A = {claro, oscuro}.

Sean A y B dos conjuntos. Una función de A x A hacia B se llama una *operación binaria* sobre el conjunto A. Encontraremos con mayor frecuencia funciones de A x A hacia A. Diremos que una función de A x A hacia A es una operación binaria *cerrada*. En el ejemplo de la máquina vendedora, se definió una operación binaria sobre el conjunto {moneda de 100, moneda de 250}. Esta operación binaria no es una operación cerrada. En el ejemplo del color de cabello de los niños, se definió una operación binaria sobre el conjunto {claro, oscuro}. Además, esta operación binaria es cerrada. Por intuición podemos afirmar que una operación binaria especifica una manera en la cual dos elementos son "combinados" para dar origen a un tercer elemento.

---

t Para ser precisos, aquí describimos el funcionamiento de una máquina vendedora en la que sólo se depositan dos monedas, las cuales son de diez y veinticinco centavos. El funcionamiento de las máquinas vendedoras reales es menos restrictivo.

Monedas depositadas	Producto entregado
(10¢, 10¢)	goma de mascar
(10¢, 25¢)	caramelo
(25¢, 10¢)	caramelo
(25¢, 25¢)	cigarros

a)

1ª moneda depositada \ 2ª moneda depositada	10¢	25¢
10¢	goma de mascar	caramelo
25¢	caramelo	cigarros

Producto entregado

b)

Padre \ Madre	claro	oscuro
claro	claro	oscuro
oscuro	oscuro	oscuro

Hijo

Figura 11.2

Una operación binaria puede describirse mediante una notación funcional. Esto es, consideremos la función de $A \times A$ hacia A llamada f. Donde $f(a_1, a_2)$[†] denotará la imagen del par ordenado (a_1, a_2) en $A \times A$. Una manera alternativa de escribir $f(a_1, a_2)$ es escribirla como $a_1 f a_2$.[‡] En este caso, en lugar de las acostumbradas letras para los nombres de funciones, usamos "símbolos de operadores" como $\star$, $*$, $+$, $\cdot$, $\square$, $\oplus$, ... como los nombres de las operaciones binarias sobre un conjunto. Así, podemos escribir

$$\star(a_1, a_2) \qquad +(\text{moneda de } 10¢, \text{ moneda de } 25¢) \qquad \square(\text{claro, oscuro})$$

o

$$a_1 \star a_2 \qquad \text{moneda de } 10¢ + \text{moneda de } 25¢ \qquad \text{claro} \ \square \ \text{oscuro}$$

La definición de operación binaria puede extenderse. Una operación ternaria sobre un conjunto *A* es una función de *(A* x *A)* x *A* hacia *B* para algún conjunto *B*, y una operación w-aria sobre un conjunto *A* es una función de A''^1 hacia *B* para algún conjunto *B}*

Un conjunto asociado con un número de operaciones sobre el conjunto recibe el nombre de *sistema algebraico.* Usaremos la notación $(A, \star, *, \square)$ para un sistema algebraico, donde *A* es un conjunto, y $\star$, $*$ y $\square$ son operaciones sobre *A*. Nuestros ejemplos de la máquina vendedora y del color del cabello de los niños son ejemplos de sistemas algebraicos con una operación. Consideremos un café de autoservicio con dos máquinas vendedoras. Podemos describir los productos que uno puede comprar mediante el sistema algebraico ({moneda de diez centavos, moneda de veinticinco centavos}., $\star$, $*$), Donde las operaciones binarias $\star$ y $*$

---

t Para ser precisos, deberíamos escribir/[(a,, a_2)]. No obstante, no existe confusión en nuestra notación. % *f(a¡, a₂)* se conoce como la notación *prefija,* y *a¡fa₂* como la notación *infijo.* § Escribimos *A'''* para denotar *((Ax A)xA)x … xA.*

★	10¢	25¢
10¢	goma de mascar	caramelo
25¢	caramelo	cigarros

*	10¢	25¢
10¢	papas	galletas
25¢	galletas	emparedado

Figura 11.3

se describen en la figura 11.3. Consideremos el conjunto de los números naturales *N* asociado con las operaciones usuales de adición y multiplicación de enteros, $+$ y $\cdot$.Es obvio que $(N, +, \cdot)$ es un sistema algebraico con dos operaciones. Asimismo, sean $\square$ una operación binaria sobre JV tal que $\square(a, b)$ es igual a 0 o 1 dependiendo de si la suma de *a* y *b* es par o impar,

y A una operación ternaria sobre //tal que $\Delta(a, b, c)$ es igual al máximo de a, b y c. $(N, \square, \Delta)$ forman un sistema algebraico con dos operaciones.

11.2 GRUPOS

Sea • una operación binaria sobre el conjunto *A.* Diremos que la operación • es *asociativa* si

$$(a \star b) \star c = a \star (b \star c)$$

para todo a, b y c en A.† Sean *A* un conjunto de personas y A una operación binaria tal que *a* A *b* es igual al más alto de *a* y *b* (supongamos que no hay dos personas de la misma estatura en *A).* Observemos que A es una operación asociativa. Por otro lado, sean *N* el conjunto de todos los números naturales y $\square$ una operación tal que $a \square b$ es igual al valor de $a^2 + b$.El lector verificará que la operación $\square$ no es asociativa. Afirmamos por intuición que cuando se tiene que llevar a cabo un cierto número de veces una operación asociativa, el orden en el cual se lleven a cabo las operaciones no es importante.

Sea $(A, \star)$ un sistema algebraico donde $\star$ es una operación binaria sobre *A.* Diremos que $(A, \star)$ es un *semigrupo* si se satisfacen las siguientes condiciones:

1. * es una operación cerrada.‡
2. • es una operación asociativa.

Sea *A* el conjunto de todos los enteros pares positivos $\{2, 4, 6, \ldots\}$ y + la operación ordinaria de adición de enteros. Puesto que + es una operación cerrada sobre *A y* también es una operación asociativa, entonces $(A, +)$ es un semigrupo. Sea *S* un alfabeto finito.

---

† Se tiene que cuando • es una operación asociativa, podemos escribir (a • *b) * c* como *a* • *b* • *c* sin ninguna confusión posible.

‡ Observemos que es posible tener una operación asociativa que no es cerrada. Considérese el sistema algebraico ({1, 2, 3}, +), donde + es la operación ordinaria de adición de enteros.

★	α	β	γ	δ
α	δ	α	β	γ
β	α	β	γ	δ
γ	α	β	γ	γ
δ	α	β	γ	δ

a)

★	α	β	γ	δ
α	α	β	δ	γ
β	β	α	γ	δ
γ	γ	δ	α	β
δ	δ	δ	β	γ

b)

Figura 11.4

Denotemos por/í al conjunto de todas las cadenas no vacías de letras de *S* (por ejemplo, sea $S = \{\alpha, \beta, \gamma\}$. Tenemos que $A = \{\alpha, \beta, \gamma, \alpha\alpha, \alpha\beta, \alpha\gamma, \ldots, \alpha\alpha\alpha, \alpha\alpha\beta, \ldots\}$).Sea • una operación binaria sobre *A* tal que para dos cadenas cualesquiera *a y b* en *A, a . b* da lugar a una cadena que es la concatenación de las cadenas a y b (por ejemplo, $\alpha\alpha \cdot \alpha\gamma\beta = \alpha\alpha\alpha\gamma\beta$). Observemos que $(A, \cdot)$ es un semigrupo.

Sea $(A, \star)$ un sistema algebraico donde • es una operación binaria sobre *A*. Diremos que un elemento en *A*, *e*, es una *identidad por la izquierda* si para todo *x* en A, $e \star x = x$. Por ejemplo, para el sistema algebraico mostrado en la figura 11.4a, tanto (3 como 8 son identidades por la izquierda. Diremos que un elemento en *A*, *e*, es una *identidad por la derecha* si para todo *x* en A, $x \star e = x$. Por ejemplo, para el sistema algebraico mostrado en la figura 11.46, a es una identidad por la derecha. Diremos que un elemento en *A* es una *identidad si* éste es tanto una identidad por la izquierda como una identidad por la derecha.

Supongamos que e_x es una identidad por la izquierda y e_2 es una identidad por la derecha de un sistema algebraico $(A, \star)$. Puesto que e_x es una identidad por la izquierda, $e_1 \star e_2 = e_2$; e_2 es una identidad por la derecha, $e_1 \star e_2 = e_1$. Así, tenemos que $e_1 = e_2$. Concluimos que si *e* es una identidad por la izquierda, entonces ya sea que *e* también sea una identidad por la derecha o no exista identidad por la derecha. De modo similar, si *e* es una identidad por la derecha, entonces ya sea que *e* también sea una identidad por la izquierda, o no exista identidad por la izquierda. *De esto se obtiene que, con respecto a una operación binaria, existe a lo más una identidad.*

De modo intuitivo, afirmamos que una identidad es un elemento "neutral" en el sentido de que, cuando se "combina" con otro elemento, su efecto sobre el resultado es nulo. Por ejemplo, sea $(A, \star)$ un sistema algebraico, donde *A* es un conjunto de luces de colores y • es una operación binaria tal que *a* • *b* es la luz de color resultante cuando la luz *a* se combina con la luz *b*. Es obvio que, la luz blanca es la identidad del sistema algebraico. Otro ejemplo, sea *(N,* +) un sistema algebraico, donde *N* es el conjunto de los números naturales y + es la operación ordinaria de adición de enteros. Es evidente que 0 es la identidad del sistema algebraico.

Sea $(A, \star)$ un sistema algebraico, donde • es una operación binaria sobre *A*. Diremos que $(A, \star)$ es un *monoide* si se satisfacen las siguientes condiciones:

1. * es una operación cerrada.
2. * es una operación asociativa.
3. Existe una identidad.

Por ejemplo, sea *A* un conjunto de personas de diferentes estaturas. Sea Δ la operación binaria tal que $a \Delta b$ es igual al más alto de a y b. Observamos que (A, Δ) es un monoide donde la identidad es la persona más baja en *A*.

$\star$	α	β	γ	δ
α	α	β	γ	δ
β	β	δ	α	γ
γ	γ	β	β	α
δ	δ	α	γ	δ

Figura 11.5

Sea $(A, \star)$ un sistema algebraico con una identidad e. Sea a un elemento en A. Diremos que un elemento b es un *inverso por la izquierda* de a si $b \star a = e$. Diremos que un elemento b es un *inverso por la derecha* de a si $a \star b = e$. Por ejemplo, para el sistema algebraico de la figural 1.5, a es una identidad, así β es un inverso por la izquierda de γ, y δ es un inverso por la derecha de *y*. Diremos que un elemento b es un *inverso* de a si éste es tanto un inverso por la izquierda como un inverso por la derecha de a. Es evidente que si b es un inverso de a, a también es un inverso de b. Por intuición afirmamos que un inverso de un elemento "cancela" el efecto del elemento cuando éstos se "combinan". Por ejemplo, sea $(A, \star)$ un sistema algebraico, donde A es un conjunto de sustancias químicas: ácidos, bases y agua, y $\star$ es una operación binaria dada como el producto de la combinación de dos sustancias químicas. En este caso, el agua puede considerarse como una sustancia química neutra, y el inverso de un ácido es una base, si su combinación produce agua.

Sea $(A, \star)$ un sistema algebraico, donde $\star$ es una operación binaria. Decimos que $(A, \star)$ es un *grupo* si se satisfacen las siguientes condiciones:

1. $\star$ es una operación cerrada.
2. $\star$ es una operación asociativa.
3. Existe una identidad.
4. Todo elemento de A tiene un inverso por la izquierda.

Observemos que *debido a la asociatividad, un inverso por la izquierda de un elemento también es un inverso por la derecha del elemento en un grupo.* Sea b el inverso por la izquierda de a y c el inverso por la izquierda de b. Sea e la identidad.[†] Dado que

$$(b \star a) \star b = e \star b = b$$

tenemos que

$$c \star ((b \star a) \star b) = c \star b = e$$

y a partir de

$$\begin{aligned} c \star ((b \star a) \star b) &= ((c \star b) \star a) \star b \\ &= (e \star a) \star b \\ &= a \star b \end{aligned}$$

† Recordamos al lector que la identidad es única.

$\oplus$	PAR	IMPAR
PAR	PAR	IMPAR
IMPAR	IMPAR	PAR

Figura 11.6

tenemos que

$$a \star b = e$$

Luego, *b* también es un inverso por la derecha de *a*. En consecuencia, nos referiremos a los inversos de los elementos sin distinguir inversos por la izquierda de inversos por la derecha. Además, observamos que la *asociatividad también implica la unicidad del inverso de un ele*... mos que tanto *b* como *c* son inversos de *a*. Esto es,

$$b \star a = e \text{ y } c \star a = e$$

De

$$(b \star a) \star b = (c \star a) \star b$$

$$b \star (a \star b) = c \star (a \star b)$$

o o

$$b = c$$

De aquí en adelante, usaremos a^{-1} para denotar el inverso de *a*.

Antes nos hemos encontrado con muchos ejemplos de grupos. Consideremos el sistema algebraico (/, +), donde /es el conjunto de todos los enteros y + es la operación ordinaria de adición de enteros. Es claro que (/, +) es un grupo que tiene al 0 como la identidad y el inverso de *n* es -*n*. Sean *G* = {PAR, IMPAR} y © una operación binaria definida como en la figura 11.6. Podemos verificar de inmediato que *(G, ®)* es un grupo donde PAR es la identidad y tanto PAR como IMPAR son sus propios inversos. Consideremos la rotación de figuras geométricas en un plano. Sea *R* = {0°, 60°, 120°, 180°, 240°, 300°} las seis posibles maneras de rotar figuras geométricas trazadas sobre un plano, a saber, rotar las figuras 0°, 60°, 120°,..., 300°. Sea • una operación binaria tal que para a y b en R, $a \star b$ es la rotación angular total correspondiente a la rotación sucesiva de *a* seguida de b. $(R, \star)$ es un grupo con 0° como la identidad, el inverso de 60° es 300°, el inverso de 180° es él mismo, etcétera. Sea $Z_n = \{0, 1, 2, \dots, n-1\}$. Sea $\oplus$ una operación binaria sobre Z_n tal que para a y b en Z_n se tiene que

$$a \oplus b = \begin{cases} a + b & \text{si } a + b < n \\ a + b - n & \text{si } a + b \geqq n \end{cases}$$

Puede verificarse rápidamente que $(Z_n, \oplus)$ es un grupo para cualquier n. $(Z_n, \oplus)$ se conoce comúnmente como el *grupo de enteros módulo n.*

Sea $\star$ una operación binaria sobre *A*. Diremos que la operación $\star$ es *conmutativa* si

$$a \star b = b \star a$$

para todo a, b en A. Sea *A* un conjunto de personas y Δ una operación binaria tal que $a \Delta b$ es igual al más alto entre *a* y *b*, y es igual a *a* si *a y b* tienen la misma estatura. Es obvio, Δ no es una operación conmutativa (¿continúa Δ siendo una operación asociativa?).

Se dice que un grupo $(A, \star)$ es un grupo *conmutativo* o un grupo *abeliano*,[†] si $\star$ es una operación conmutativa. Por ejemplo, $(Z_n, \oplus)$ es un grupo conmutativo.

Se dice que un grupo $(A, \star)$ es *finito* si *A* es un conjunto finito, e *infinito* si *A* es un conjunto infinito. El tamaño de *A* es conocido como el *orden* del grupo.

11.3 SUBGRUPOS

Sea $(A, \star)$ un sistema algebraico y *B* un subconjunto de *A*. Diremos que el sistema algebraico $(B, \star)$ es un subsistema de $(A, \star)$. La noción de subsistema es una muy natural. Supongamos que $(A, \star)$ es un sistema algebraico que describe la interacción de un conjunto de partículas atómicas. Si estamos interesados en la interacción de algunas de las partículas, sólo podemos considerar un subsistema de $(A, \star)$. Sea $(N, +)$. , un sistema algebraico que describe la adición de números naturales. *(E,* +) es un subsistema de $(N, +)$ si E es el conjunto de todos los números pares. De modo similar, consideremos el ejemplo de la rotación de figuras geométricas en un plano. Observemos que $(\{0°, 120°, 240°\}, \star)$ es un subsistema del sistema algebraico $(\{0°, 60°, 120°, 180°, 240°, 300°\}, \star)$. Al igual que lo es $(\{0°, 180°\}, \star)$.

Sea $(A, \star)$ un grupo, y *B* un subconjunto de *A*. Diremos que $(B, \star)$ es un *subgrupo* de A si $(B, \star)$ es a su vez un grupo. Supongamos que queremos verificar si $(B, \star)$ es un subgrupo para un subconjunto *B* de *A*. Observemos que:

1. Debemos verificar si • es una operación cerrada sobre *B.*
2. Se sabe que • es una operación asociativa.
3. Debido a que sólo existe un elemento *e* en *A* tal que $e \star x = x \star e = x$ para todo x en A, debemos verificar que *e* está en *B.* En otras palabras, la identidad de $(A, \star)$ debe estar en *B* como la identidad de $(B, \star)$.
4. Puesto que el inverso de cualquier elemento en *A* es único, para cualquier elemento *b* en *B,* debemos verificar que su inverso también está en *B.*

Por ejemplo, sea $(I, +)$ un sistema algebraico, donde / es el conjunto de todos los enteros y + es la operación ordinaria de adición de enteros. Resulta claro que (/, +) es un grupo. Además, $(E, +)$ es un subgrupo, donde *E* es el conjunto de todos los enteros pares. De igual

En honor del matemático noruego N. H. Abel (1802-1829).

manera, en el ejemplo de la rotación de figuras geométricas en el plano, tanto $(\{0°, 120°, 240°\}, \star)$ como $(\{0°, 180°\}, \star)$ son subgrupos del grupo $(\{0°, 60°, 120°, 180°, 240°, 300°\}, \star)$.

Queremos demostrar que:

Teorema 11.1 Sea $(A, \star)$ un grupo y B un subconjunto de A. Si B es un conjunto finito, entonces $(B, \star)$ es un subgrupo de $(A, \star)$ si $\star$ es una operación cerrada sobre B.

DEMOSTRACIÓN Sea a un elemento en B. Si $\star$ es una operación cerrada sobre B, los elementos $a, a^2, a^3, \ldots$ están todos en B.† Debido a que B es un conjunto finito, de acuerdo con el principio del palomar, debemos tener que $a^i = a^j$ para algún i y j, $i < j$. Esto es, $a^i = a^i \star a^{j-i}$. Entonces a^{j-i} es la identidad de $(A, \star)$, y está incluida en el subconjunto B. Si $j - i > 1$, de acuerdo con $a^{j-i} = a \star a^{j-i-1}$, podemos concluir que a^{j-i-1} es el inverso de a, y está incluido en el subconjunto B. Si $j - i = 1$, tenemos que $a^i = a^i \star a$. Así, a debe ser el elemento identidad y su propio inverso. En consecuencia, el que $\star$ sea una operación cerrada sobre B garantiza que $(B, \star)$ es un subgrupo. ❐

11.4 GENERADORES Y EVALUACIÓN DE POTENCIAS

Sea $(A, \star)$ un sistema algebraico en el cual A es un conjunto de colores y la operación binaria $\star$ proporciona el color que origina la combinación de dos colores, por ejemplo, rojo $\star$ amarillo = naranja. Supongamos que se nos da un subconjunto de colores en A; podríamos querer conocer todos los colores que obtendríamos si utilizáramos todas las posibles combinaciones de los colores suministrados. De modo similar, consideremos el grupo $(\{0°, 60°, 120°, 180°, 240°, 300°\}, \star)$ que describe la rotación de figuras geométricas en el plano. Supongamos que sólo podemos rotar las figuras 120° en cada ocasión. Rotaciones sucesivas de 120° darán lugar a las rotaciones {0°, 120°, 240°}. Por otro lado, supongamos que sólo podemos rotar las figuras 60° en cada ocasión. Rotaciones sucesivas de 60° darán lugar a todas las rotaciones en (0°, 60°, 120° 180°, 240°, 300°}.

Sean $(A, \star)$ un sistema algebraico, donde • es una operación cerrada, y $B = \{a_1, a_2, \ldots\}$ un subconjunto de A. Sea B_1 el subconjunto de A el cual contiene a B como también a todos los elementos $a_i \star a_j$ para a_i y a_j en B. B_1 recibe el nombre de conjunto *generado directamente* por B. De modo similar, sean B_2 el conjunto generado directamente por $B_1, \ldots$, y B_{i+1} el conjunto generado directamente por B_i. Sea B^* la unión de $B, B_1, B_2, \ldots$ El sistema algebraico $(B^*, \star)$ recibe el nombre de *subsistema generado* por B, y diremos que un elemento es generado por B si éste se encuentra en B^*. Observemos que $\star$ es una operación cerrada sobre B^*. Así, para un grupo $(A, \star)$, si B^* es finito, entonces $(B^*, \star)$ es un subgrupo. Si $B^* = A$, diremos que 6 es un *conjunto generador* o un *conjunto de generadores* del sistema algebraico $(A, \star)$. En el ejemplo sobre la combinación de colores, un conjunto generador es un subconjunto del conjunto de colores cuyas combinaciones darán lugar a todos los colores

† Usamos a^m para denotar $\underbrace{a \star a \star a \cdots \star a}_{m \text{ veces}}$.

$\star$	α	β	γ	δ
α	α	β	γ	δ
β	β	γ	δ	α
γ	γ	δ	α	β
δ	δ	α	β	γ

a)

$\star$	α	β	γ	δ
α	α	β	γ	δ
β	β	α	δ	γ
γ	γ	δ	α	β
δ	δ	γ	β	α

b)

Figura 11.7

en el conjunto original. En el ejemplo sobre la rotación de figuras geométricas en el plano, {60°} es un conjunto generador, al igual que lo es {120°, 180°}.

Un grupo que tiene un conjunto generador consistente en un único elemento se conoce como un *grupo cíclico*. La figura 11 *.la* muestra un grupo cíclico para él cual $\{\beta\}$ es un conjunto generador. Observemos que también $\{\delta\}$ es un conjunto generador. Para el ejemplo de la rotación de figuras geométricas en el plano, el grupo ({0°, 60°, 120°, 180°, 240°, $300°\}, \star)$ también es un grupo cíclico. El lector verificará que el grupo de la figura 11 *.Ib* no es un grupo cíclico.

Sean $(A, \star)$ un grupo cíclico y $\{a\}$ un conjunto generador de $(A, \star)$. Los elementos de *A* pueden expresarse como $a, a^2, a^3, \ldots$ Puesto que, a causa de la asociatividad, $a^i \star a^j = a^j \star a^i = a^{i+j}$, se obtiene de inmediato que *cualquier grupo cíclico es conmutativo.*

Nos desviaremos por un momento para presentar un interesante problema relacionado con el concepto de conjuntos generadores de sistemas algebraicos. Sea *B* un conjunto generador de un sistema algebraico $(A, \star)$. Para un elemento *a* en *A*, podríamos desear conocer las diferentes maneras de generar al elemento *a*. Generar al elemento *a* significa obtener *a* a través de combinaciones sucesivas de los elementos en el conjunto generador. Una manera de generar *a* puede especificarse mediante una sucesión de elementos en *A*

$$a_1 \quad a_2 \quad a_3 \quad \cdots \quad a_r$$

tal que $a_r = a$, y cada a_i, $1 \leqq i \leqq r$, puede expresarse como $a_j \star a_k$, donde a_j y a_k son ya sea elementos de *B* o elementos que preceden a *a*, en la sucesión. Puesto que una sucesión de *r* elementos corresponde a la generación del elemento *en* aplicaciones de la operación • a elementos en el conjunto generador y elementos que ya han sido generados, será interesante determinar sucesiones cortas que generen un elemento dado.

Nuestro problema está motivado por aquel problema de encontrar procedimientos eficientes para evaluar la potencia *x"* para una *x* dada y un entero positivo *n.* Consideremos el sistema algebraico $(I, +)$, donde / es el conjunto de todos los enteros positivos y + es la operación ordinaria de adición de enteros. Resulta claro que $B = \{1\}$ es un conjunto generador del sistema. Para un entero dado *n*, nos gustaría conocer las diferentes maneras con las cuales puede generarse. Por ejemplo, las siguientes sucesiones muestran algunas de las maneras de generar al entero 9:

2	3	4	5	6	7	8	9
2	3	4	5	9			
2	4	8	9				

En la literatura, una sucesión de elementos de / que conduce a la generación de un entero n se conoce como una *cadena de adición* para n. La conexión entre una cadena de adición para n y un procedimiento para evaluar x^n para un valor de x dado resulta obvia una vez que recordamos que $x^j \cdot x^k = x^{j+k}$.

El problema de determinar una cadena de adición mínima para un entero dado n es muy interesante y ha sido estudiado extensamente. A pesar de ello, un análisis minucioso de este problema se encuentra fuera del alcance de este libro. A manera de ilustración, presentamos aquí dos procedimientos simples para determinar cadenas de adición para n.†

Si n puede factorizarse *zoxnopq*, podemos determinar primero cadenas de adición para p y q y entonces combinarlas para obtener una cadena de adición para pq. Sean

$$p_1 \quad p_2 \quad \cdots \quad p_{i-1} \quad p$$

y

$$q_1 \quad q_2 \quad \cdots \quad q_{j-1} \quad q$$

las cadenas de adición para/? y q. Resulta claro,

$$q_1 \quad q_2 \quad \cdots \quad q_{j-1} \quad q \quad p_1q \quad \cdots \quad p_{i-1}q \quad pq$$

es una cadena de adición para n.

Consideremos el ejemplo de $n = 45$, el cual puede escribirse como 5×9. Puesto que

$$2 \quad 3 \quad 5$$

es una cadena de adición para 5, y

$$2 \quad 4 \quad 8 \quad 9$$

es una cadena de adición para 9, obtenemos

$$2 \quad 4 \quad 8 \quad 9 \quad 18 \quad 27 \quad 45$$

como una cadena de adición para 45. Alternativamente, dado que

$$9 = 3 \times 3$$

y como

$$2 \quad 3$$

es una cadena de adición para 3, obtenemos

$$2 \quad 3 \quad 6 \quad 9$$

como una cadena de adición para 9, y

$$2 \quad 3 \quad 6 \quad 9 \quad 18 \quad 27 \quad 45$$

como una cadena de adición para 45.

El segundo procedimiento es recursivo. Si n es un número par, podemos determinar una cadena de adición para $n/2$, y entonces agregar $n/2$ a $n/2$ para obtener n. Así, si

$$a_1 \quad a_2 \quad \cdots \quad n/2$$

† Véase también el problema 8.11.

es una cadena de adición para *nll*, entonces

$$a_1 \quad a_2 \quad \cdots \quad n/2 \quad n$$

es una cadena de adición para *n*. Si *n* es un número impar, podemos determinar una cadena de adición para $(n-1)/2$, agregar $(n-1)/2$ a $(n-1)/2$ para obtener $n-1$, y entonces agregar 1 a *n* - 1 para obtener *n*. Así, si

$$a_1 \quad a_2 \quad \cdots \quad (n-1)/2$$

es una cadena de adición para $(n-1)/2$, entonces

$$a_1 \quad a_2 \quad \cdots \quad (n-1)/2 \quad n-1 \quad n$$

es una cadena de adición para *n*. Ahora, este procedimiento puede aplicarse para determinar una cadena de adición para $n/2$ o para $(n-1)/2$. Por ejemplo,

$$2 \quad 4 \quad 5 \quad 10 \quad 11 \quad 22 \quad 44 \quad 45$$

es una cadena de adición para el entero 45, determinada mediante este procedimiento.

11.5 CONJUNTOS COCIENTE Y EL TEOREMA DE IAGRANGE

Consideremos una extensión del ejemplo de la máquina vendedora de la sección 11.1 de manera que en cada compra se depositen dos monedas del conjunto {moneda de 5¢, moneda de 10¢, moneda de 25¢, moneda de 50¢, moneda de $1}. Si ya hemos depositado una moneda de 25¢ en la máquina, podríamos desear conocer los productos que recibiremos si la segunda moneda que depositaremos es una moneda de 10¢, una moneda de 25¢ o una moneda de 50¢. . De igual manera, consideremos el ejemplo de la rotación de figuras geométricas en el plano. Supongamos que una rotación inicial de 0°, 120° o 240° es seguida por una rotación de 60°. Queremos conocer las posibles rotaciones angulares totales. Sea $(A, \star)$ un sistema algebraico, donde $\star$ es una operación binaria. Sea *a* un elemento en *A*, y *H* un subconjunto de *A*. El *conjunto cociente por la izquierda de H respecto de a*, el cual denotamos por $a \star H$, es el conjunto de elementos $\{a \star x | x \in H\}$. De modo similar, el *conjunto cociente por la derecha de H respecto de a*, al cual denotamos por $H \star a$, es el conjunto de elementos $\{x \star a | x \in H\}$. Resulta claro que en el ejemplo de la máquina vendedora, queremos determinar el conjunto cociente por la izquierda del conjunto {moneda de 10¢, moneda de 25¢, moneda de 50¢} respecto a una moneda de 25¢, y en el ejemplo de la rotación de figuras geométricas, queremos determinar el conjunto cociente por la derecha del conjunto {0°, 120°, 240°} respecto de 60°.

Podemos decir mucho más acerca de conjuntos cociente cuando nos restringimos a conjuntos cociente en grupos. Sea $(A, \star)$ un grupo y $(H, \star)$ un subgrupo de $(A, \star)$. Tenemos que:

teorema 11.2 Sean $a \star H$ y $b \star H$ _ dos conjuntos cociente de H. Bien $a \star H$ y $b \star H$ son disjuntos o bien son idénticos.

DEMOSTRACIÓN Supongamos que $a \star H$ y $b \star H$ no son disjuntos, y tienen a/como elemento común. Esto es, existen h_1 y h_2 en H tales que $f = a \star h_1 = b \star h_2$. Podemos escribir $a = b \star h_2 \star h_1^{-1}$. Para un elemento cualquiera x en $a \star H$, ya que $x = a \star h_3$ para algún h_3 en H, tenemos que $x = b \star h_2 \star h_1^{-1} \star h_3$ el cual es un elemento de $b * H$ debido a que $h_2 \star h_1^{-1} \star h_3$ es un elemento en H. En forma similar podemos demostrar que cualquier elemento en $b \bullet H$ también es un elemento en $a \bullet H$. Así concluimos que los dos conjuntos $a \star H$ y $b \star H$ son iguales,[†]

□

Sean $(A, \star)$ un grupo, y $(H, \star)$ un subgrupo de $(A, \star)$ Debido a que $(A, \star)$ es un grupo, para cualquier a en A y cualesquiera distintos h_1 y h_2 en H, $a \star h_1 \neq a \star h_2$. Se tiene que el tamaño de cualquier conjunto cociente de //es el mismo que el de H. Además, debido a que H contiene a la identidad del grupo, si calculamos todos los conjuntos cociente por la izquierda (por la derecha) de //, habremos usado exhaustivamente todos los elementos en A. En consecuencia, podemos concluir que los conjuntos cociente por la izquierda de H forman una partición de A, en la cual todos los bloques son del mismo tamaño. Así, el tamaño de A es igual al número de diferentes conjuntos cociente por la izquierda de H multiplicado por el tamaño de H. En otras palabras tenemos:

Teorema 11.3 (Lagranee) El orden de cualquier subgrupo de un grupo finito divide al orden del grupo.

Existen algunas consecuencias inmediatas del teorema de Lagrange. Primero, observamos que un grupo de primer orden no tiene un subgrupo no-trivial.[‡] De esto se tiene que un grupo de primer orden debe ser cíclico, y que cualquier conjunto de un solo elemento distinto de la identidad es un conjunto generador.

*11.6 GRUPOS DE PERMUTACIONES Y TEOREMA DE BURNSIDE

En esta sección estudiaremos una clase importante de grupos. Una función uno a uno de un conjunto S sobre él mismo se conoce como una *permutación* del conjunto S. Usamos la notación $\begin{pmatrix} abcd \\ bdca \end{pmatrix}$ para la permutación del conjunto $\{a, b, c, d\}$ que mapea a hacia b, b hacia d, c hacia c y ¿/hacia a; esto es, en la fila superior los elementos en el conjunto están escritos en un orden arbitrario, y en la fila inferior la imagen de un elemento será escrita bajo el elemento mismo.

Para un conjunto de n elementos, S, denotemos *por A* al conjunto de las «! permutaciones de S. Definimos una operación binaria ° sobre A como la composición de dos funciones (véase problema 4.32). Observamos que *la operación binaria ° es una operación cerrada sobre A.Sean* π_1 y π_2 dos permutaciones del conjuntoPara $S = \{a, b, c, \ldots, x, y, z\}$. demostrar que $\pi_1 \circ \pi_2$ es también una permutación del conjunto S, sólo tenemos que demostrar que no hay dos elementos en Sque sean mapeados hacia el mismo elemento por $\pi_1 \circ \pi_2$. Supongamos que π_2 mapea al elemento a hacia b y que TI, mapea al elemento b hacia c. $\pi_1 \circ \pi_2$ mapeará

---

t Queremos recordarle al lector que esto no significa que $a \star h = b \bullet h$ para todo h en *fi*.

% Diremos que un subgrupo es trivial si contiene todos los elementos del grupo o solamente a la identidad.

entonces al elemento *a* hacia c. Sea x cualquier elemento distinto de *a*. Puesto que *%*$_2$ es una permutación del conjunto *S*, *n*$_2$ mapea *x* hacia un elemento que es distinto de *b*, digamos *y*. De modo similar, it, mapea *y* hacia un elemento que es distinto de *c*, digamos z. Así, $\pi_1 \circ \pi_2$ mapea *x* hacia *z*. Luego concluimos que $\pi_1 \circ \pi_2$ siempre mapeará dos elementos distintos (por ejemplo, a y x) hacia dos elementos distintos (por ejemplo, c y z) y es, por tanto, una permutación del conjunto *S*. Por ejemplo, si

$$\pi_1 = \begin{pmatrix} abcd \\ adbc \end{pmatrix} \qquad \pi_2 = \begin{pmatrix} abcd \\ bacd \end{pmatrix}$$

tenemos que

$$\pi_1 \circ \pi_2 = \begin{pmatrix} abcd \\ dabc \end{pmatrix}$$

También observamos que *la operación binaria o es asociativa*. Esto es, para las permutaciones cualesquiera π_1, π_2 y π_3 de un conjunto, tenemos que $(\pi_1 \circ \pi_2) \circ \pi_3 = \pi_1 \circ (\pi_2 \circ \pi_3)$. Este hecho puede verse como sigue: supongamos que 7t$_3$ mapea *a* hacia *b*, *n*$_2$ mapea *b* hacia *c*, y TI, mapea c hacia *d*. Debido a que $\pi_1 \circ \pi_2$ mapea b hacia d, $(\pi_1 \circ \pi_2) \circ \pi_3$ mapea *a* hacia *d*. De modo similar, ya que $\pi_2 \circ \pi_3$ mapea *a* haciamapea c, $\pi_1 \circ (\pi_2 \circ \pi_3)$ *a* hacia *d*. Por ejemplo, si

$$\pi_1 = \begin{pmatrix} abcd \\ adbc \end{pmatrix} \qquad \pi_2 = \begin{pmatrix} abcd \\ bacd \end{pmatrix} \qquad \pi_3 = \begin{pmatrix} abcd \\ bdac \end{pmatrix}$$

Entonces

$$(\pi_1 \circ \pi_2) \circ \pi_3 = \left[\begin{pmatrix} abcd \\ adbc \end{pmatrix}\begin{pmatrix} abcd \\ bacd \end{pmatrix}\right]\begin{pmatrix} abcd \\ bdac \end{pmatrix} = \begin{pmatrix} abcd \\ dabc \end{pmatrix}\begin{pmatrix} abcd \\ bdac \end{pmatrix} = \begin{pmatrix} abcd \\ acdb \end{pmatrix}$$

y

$$\pi_1 \circ (\pi_2 \circ \pi_3) = \begin{pmatrix} abcd \\ adbc \end{pmatrix}\left[\begin{pmatrix} abcd \\ bacd \end{pmatrix}\begin{pmatrix} abcd \\ bdac \end{pmatrix}\right] = \begin{pmatrix} abcd \\ adbc \end{pmatrix}\begin{pmatrix} abcd \\ adbc \end{pmatrix} = \begin{pmatrix} abcd \\ acdb \end{pmatrix}$$

De esto se tiene que $(A, \circ)$ es un grupo en el cual la permutación que mapea cualquier elemento de *S* hacia él mismo es la identidad, y el inverso de una permutación *n* es uno que mapea *n(a)* hacia *a* para todo *a* en *S*. Por ejemplo, para $S = \{a, b, c, d\}$, la identidad de $(A, \circ)$ es $\begin{pmatrix} abcd \\ abcd \end{pmatrix}$, el inverso de $\begin{pmatrix} abcd \\ bcad \end{pmatrix}$ es $\begin{pmatrix} abcd \\ cabd \end{pmatrix}$. Un subgrupo de $(A, \circ)$ usualmente es referido como un *grupo de permutación* del conjunto *S*.

Sea $(G, \circ)$ un grupo de permutación del conjunto $S = \{a, b, \ldots\}$. Una relación binaria sobre el conjunto *S*, es llamada la *relación binaria inducida por* (G, o), está definida de tal manera que el elemento *a* está relacionado al elemento *b* si y sólo si existe una permutación en *G* que mapea *a* hacia *b*. Por ejemplo, si

$$G = \left\{\begin{pmatrix} abcd \\ abcd \end{pmatrix}, \begin{pmatrix} abcd \\ bacd \end{pmatrix}, \begin{pmatrix} abcd \\ abdc \end{pmatrix}, \begin{pmatrix} abcd \\ badc \end{pmatrix}\right\}$$

La relación binaria inducida por (G, o) se muestra en la figura 11.8. Observemos que *la relación binaria sobre S inducida por un grupo de permutación* (G, °) *es una relación de*

	a	b	c	d
a	√	√		
b	√	√		
c			√	√
d			√	√

Figura 11.8

equivalencia. Debido a que la permutación identidad está en *G,* cualquier elemento en *S* está relacionado a él mismo en la relación binaria sobre *S* inducida por $(G, \circ)$. Entonces se satisface la ley de reflexividad. Si existe una permutación 7tj en *G* que mapea *a* hacia *b,* el inverso de 7i,, el cual también está en *G,* mapeará *b* hacia *a.* Por tanto, la relación binaria sobre *S* inducida por $(G, \circ)$ satisface la ley de simetría. Si existe una permutación n_l mapeando *a* hacia *b,* y una permutación n_2 mapeando *b* hacia *c,* la permutación $\pi_2 \circ \pi_1$, la cual también está en *G,* mapeará *a* hacia *c.* Así, la relación binaria sobre *S* inducida por $(G, \circ)$ satisface la ley de transitividad.

Ahora estamos preparados para demostrar un resultado que obtuvo Burnside, conocido como el *teorema de Burnside.* Dado un conjunto *S* y un grupo de permutación $(G, \circ)$ de S, deseamos encontrar el número de clases de equivalencia en las cuales se divide *S* por la relación de equivalencia sobre *S* inducida por $(G, \circ)$. Este problema se resuelve directamente al hallar la relación de equivalencia y contar entonces el número de clases de equivalencia. Sin embargo, cuando el conjunto *S* contiene un gran número de elementos, ese conteo se vuelve demasiado tedioso. La alternativa es el teorema de Burnside para encontrar el número de clases de equivalencia mediante el conteo del número de elementos que son invariantes bajo las permutaciones en el grupo. Diremos que un elemento es un *invariante* bajo una permutación, o se le denomina como una *invariancia,* si la permutación mapea al elemento hacia él mismo.

Teorema 11.4 (Burnside)

El número de clases de equivalencia en las cuales es dividido un conjunto *S* mediante una relación de equivalencia inducida por un grupo de permutación $(G, \circ)$ de *S* está dado por

$$\frac{1}{|G|} \sum_{\pi \in G} \psi(\pi)$$

donde $\psi(\pi)$ es el número de elementos que son invariantes bajo la permutación *n.*

Para que podamos apreciar mejor el significado del teorema de Burnside, ilustraremos sus aplicaciones antes de proceder con la demostración. Sea $S = \{a, b, c, d\}$, y sea *G* el grupo de permutación que consiste en

$$\pi_1 = \begin{pmatrix} abcd \\ abcd \end{pmatrix} \qquad \pi_2 = \begin{pmatrix} abcd \\ bacd \end{pmatrix} \qquad \pi_3 = \begin{pmatrix} abcd \\ abdc \end{pmatrix} \qquad \pi_4 = \begin{pmatrix} abcd \\ badc \end{pmatrix}$$

La relación de equivalencia sobre S inducida por G se muestra en la figura 11.8. S está dividida en dos clases de equivalencia, $\{a, b\}$ y $\{c, d\}$. Para calcular el número de clases de equivalencia de acuerdo con el teorema de Burnside, observemos que puesto que $\psi(\pi_1) = 4$, $\psi(\pi_2) = 2$, $\psi(\pi_3) = 2$ y $\psi(\pi_4) = 0$, el número de clases de equivalencia es

$$\tfrac{1}{4}(4 + 2 + 2 + 0) = 2$$

DEMOSTRACIÓN Para cualquier elemento s en S, sea $\eta(s)$ el número de permutaciones bajo las cuales 5 es invariante. Entonces

$$\sum_{\pi \in G} \psi(\pi) = \sum_{s \in S} \eta(s)$$

debido a que tanto $\Sigma_{\pi \in G}\, \psi(\pi)$ como $\Sigma_{s \in S}\, \eta(s)$ cuentan el número total de invariantes bajo todas las permutaciones en G [una manera de contar las invariancias es revisar las permutaciones una por una y contar el número de invariancias bajo cada permutación. Esto da $\Sigma_{\pi \in G}\, \psi(\pi)$ como la suma total. Otra manera de contar las invariancias es avanzar conforme los elementos de 5 uno por uno y contar el número de permutaciones bajo las cuales el elemento en turno es invariante. Esto da $\Sigma_{s \in S}\, \eta(s)$ como la suma total].

Sean a y b dos elementos en S que están en la misma clase de equivalencia. Queremos demostrar que existen exactamente $\eta(a)$ permutaciones mapeando a hacia b. Puesto que a y b están en la misma clase de equivalencia, existe al menos una de dichas permutaciones a la cual denotaremos por π_x. Sea $\{\pi_1, \pi_2, \pi_3, \ldots\}$ el conjunto de las $\eta(a)$ permutaciones bajo las cuales a es invariante. Entonces, las $\eta(a)$ permutaciones en el conjunto $\{\pi_x \circ \pi_1, \pi_x \circ \pi_2, \pi_x \circ \pi_3, \ldots\}$ son permutaciones que mapean a hacia b.

Primero observemos que todas estas permutaciones son distintas porque, si $\pi_x \circ \pi_1 = \pi_x \circ \pi_2$, tenemos que

$$\pi_x^{-1} \circ (\pi_x \circ \pi_1) = \pi_x^{-1} \circ (\pi_x \circ \pi_2)$$

Esto da $\pi_1 = \pi_2$, lo cual es imposible. Segundo, vemos que ninguna otra permutación en G mapea a hacia b. Supongamos que existe una permutación π_y que mapea a hacia b. Entonces, $\pi_x^{-1} \circ \pi_y$ es una permutación que mapea a hacia a, ya que π_x^{-1} mapea b hacia a. Debido a que $\pi_x^{-1} \circ \pi_y$ es una permutación en el conjunto $\{\pi_1, \pi_2, \pi_3, \ldots\}$, $\pi_x \circ (\pi_x^{-1} \circ \pi_y) = \pi_y$ es una permutación en el conjunto $\{\pi_x \circ \pi_1, \pi_x \circ \pi_2, \pi_x \circ \pi_3, \ldots\}$. Entonces, concluimos que existen exactamente $\eta(a)$ permutaciones en G que mapean a hacia b.

Sean $a, b, c, \ldots, h$ elementos en S que están en una clase de equivalencia. Todas las permutaciones en G pueden ser categorizadas como las que mapean a hacia a, las que mapean a hacia b, las que mapean a hacia $c, \ldots,$ y las que mapean a hacia h. Puesto que hemos demostrado que existen exactamente $\eta(a)$ permutaciones en cada una de estas categorías tenemos que

$$\eta(a) = \frac{|G|}{\text{número de elementos en la clase de equivalencia que contiene a } a}$$

Mediante un argumento similar, obtenemos que

$$\begin{aligned}\eta(b) &= \eta(c) = \cdots = \eta(h)\\ &= \frac{|G|}{\text{número de elementos de la clase de equivalencia que contiene a } a}\end{aligned}$$

y, por tanto,

$$\eta(a) + \eta(b) + \eta(c) + \cdots + \eta(h) = |G|$$

De esto se tiene que, para cualquier clase de equivalencia de elementos en S,

$$\sum_{\text{todo } s \text{ en la clase de equivalencia}} \eta(s) = |G|$$

y

$$\sum_{s \in S} \eta(s) = \begin{pmatrix}\text{número de clases de equivalencia}\\ \text{en las cuales se divide } S\end{pmatrix} \times |G|$$

Luego entonces, tenemos que

Número de clases de equivalencia en las cuales se divide S

$$= \frac{1}{|G|}\sum_{s \in S} \eta(s) = \frac{1}{|G|}\sum_{\pi \in G} \psi(\pi)$$

❐

Ahora consideremos algunos ejemplos ilustrativos.

Ejemplo 11.1

Queremos encontrar el número de cadenas distintas de longitud 2 que se construyen a partir de cuentas azules y amarillas. Los dos extremos de una cadena no se marcan, y dos cadenas son, por tanto, indistinguibles ya que si se intercambian los extremos de una esto dará lugar a la otra. Sean b y y las cuentas azules y amarillas, respectivamente. Sean bb, by, yb y yy las cuatro cadenas diferentes de longitud 2 cuando la equivalencia entre cadenas no se toma en consideración. El problema es encontrar el número de clases de equivalencia en las cuales el conjunto $S = \{bb, by, yb, yy\}$ es dividido por la relación de equivalencia inducida por el grupo de permutación $(\{\pi_1, \pi_2\}, \circ)$, donde

$$\pi_1 = \begin{pmatrix} bb & by & yb & yy \\ bb & by & yb & yy \end{pmatrix} \quad \pi_2 = \begin{pmatrix} bb & by & yb & yy \\ bb & yb & by & yy \end{pmatrix}$$

La permutación π_1 indica que cualquier cadena es equivalente a ella misma, y la permutación π_2 especifica la equivalencia entre cadenas cuando los dos extremos de una cadena son intercambiados. De acuerdo con el teorema de Burnside, el número de cadenas distintas es

$$\tfrac{1}{2}(4 + 2) = 3$$

De modo similar, para el caso de cadenas distintas de longitud 3 construidas con cuentas azules y amarillas, tenemos el conjunto $S = \{bbb, bby, byb, ybb, byy, yby, yyb,$

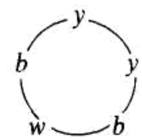
yyy} y el grupo de permutación $(G, \circ)$, $G = \{\pi_1, \pi_2\}$, donde π_1 es la permutación identidad y π_2 es la permutación que mapea una cadena hacia una que se obtiene a partir de la misma mediante el intercambio de sus extremos; por ejemplo, *bbb* es mapeada hacia *bbb*, *bby* es mapeada hacia *ybb*, *byb* es mapeada hacia *byb*, y así sucesivamente. El número de elementos que son invariantes bajo π_1 es ocho. El número de elementos que son invariantes bajo π_2 es cuatro, ya que una cadena será mapeada hacia ella misma bajo π_2 si las cuentas en los dos extremos de la cadena son del mismo color, y existen cuatro de dichas cadenas. Por consiguiente, el número de cadenas distintas es igual a

$$\tfrac{1}{2}(8 + 4) = 6$$

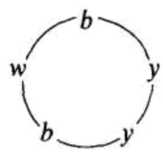
□

Ejemplo 11.2

Supongamos que queremos encontrar el número de brazaletes distintos de cinco cuentas construidos con cuentas amarillas, azules y blancas. Diremos que dos brazaletes son indistinguibles si la rotación de uno da lugar al otro. No obstante, para simplificar el problema supongamos que los brazaletes no pueden ser volteados. Sea S el conjunto de los 3^5 (= 243) brazaletes distintos cuando no se considera la equivalencia rotacional. Sea $(\{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}, \circ)$ un grupo de permutación, donde π_1 es la permutación identidad y π_2 es la permutación que mapea un brazalete hacia él mismo pero rotado en el sentido de las manecillas del reloj por una cuenta. Por ejemplo,



es mapeado hacia



De igual modo", π_3, π_4 y π_5 son permutaciones que mapean un brazalete hacia sí mismo pero rotado en el sentido de las manecillas del reloj por dos, tres y cuatro cuentas, respectivamente.

El número de elementos invariantes bajo π_1 es 243. El número de elementos invariantes bajo π_2 es tres debido a que sólo cuando las cinco cuentas en un brazalete son del mismo color su rotación por una cuenta dará lugar al mismo brazalete. De modo similar, el número de elementos que son invariantes bajo cada una de entre π_3, π_4 y π_5 es también tres. Por consiguiente, el número de brazaletes distintos es

$$\tfrac{1}{5}(243 + 3 + 3 + 3 + 3) = 51$$

El resultado sobre el número de brazaletes distintos conduce de inmediato a una prueba muy interesante de algo conocido en teoría de números como *el pequeño teorema de Fermat.* Para un número primo p, determinamos el número de brazaletes distintos de p cuentas construidos con cuentas de a colores diferentes cuando se permite la equivalencia rotacional. El número de brazaletes distintos es

$$\frac{1}{p}(a^p + a + a + \cdots + a) = \frac{1}{p}[a^p + (p-1)a]$$

Puesto que el número de brazaletes distintos es un número entero, p divide a $a^p + (p-1)a$, o p divide a $a^p - a$. Si p no divide a a, entonces p debe dividir a $a^{p-1} - 1$, lo cual es exactamente el teorema de Fermat.

❐

Ejemplo 11.3

Podemos solucionar el problema del ejemplo 3.8 mediante el teorema de Burnside. Sean S el conjunto de los 10^5 números de 5 dígitos y $(\{\pi_1, \pi_2\}, \circ)$ un grupo de permutación de S, donde π_1 es la permutación identidad, y π_2 es una permutación que mapea un número en él mismo si éste no se puede leer como un número cuando se coloca de cabeza (por ejemplo, 13765 es mapeado hacia 13765) y cuando mapea un número hacia el número obtenido por la lectura de éste al ser puesto de cabeza, siempre que se pueda leer (por ejemplo, 89166 es mapeado hacia 99168). El número de invariancias bajo Ttj es 10^5. El número de invariancias bajo n_2 es $(10^5 - 5^5) + 3 \times 5^2$ debido a que existen $10^5 - 5^5$ números que contienen uno o más dígitos de entre 2, 3, 4, 5 y 7, y en consecuencia no pueden leerse de cabeza, y debido a que existen 3×5^2 números que se leerán iguales tanto al derecho como de cabeza, por ejemplo, 16891 (el dígito central de estos números debe ser 0 o 1 u 8, el último dígito debe ser el primer dígito puesto de cabeza, y el cuarto dígito debe ser el segundo dígito puesto de cabeza). Por tanto, el número de tiras distintas que se forman es

$$\tfrac{1}{2}(10^5 + 10^5 - 5^5 + 3 \times 5^2) = 10^5 - \tfrac{1}{2} \times 5^5 + \tfrac{3}{2} \times 5^2$$

❐

11.7 CÓDIGOS Y CÓDIGOS DE GRUPO

El problema de codificación es representar mensajes distintos mediante diferentes secuencias de letras de un alfabeto dado. Por ejemplo, mensajes como "emergencia", "la ayuda está en camino", "está todo claro", etcétera, pueden ser representados por secuencias de puntos y rayas. En nuestro análisis de esta sección suponemos que el alfabeto es binario {0,1}. Una secuencia de letras de un alfabeto se conoce como una *palabra.* Un *código* es una colección de palabras que son utilizadas para representar distintos mensajes. Una palabra en un código también es denominada como *unapalabra código.* Un *código de bloque* es un código consistente de palabras que son de la misma longitud. Uno de los criterios al escoger un bloque de código para representar un conjunto de mensajes es su capacidad para corregir errores. Supongamos que una palabra código es transmitida desde su origen hacia su destino. En el curso de la transmisión, interferencias como ruido pueden ocasionar que algunos de los números uno en la palabra código sean recibidos como números cero, y algunos de los ceros como unos. En consecuencia, la palabra recibida podría no continuar siendo la palabra transmitida, y es nuestro deseo recuperar la palabra transmitida lo mejor posible. Esto es lo que entendemos por corrección de errores.

Denotemos por A al conjunto de todas las sucesiones binarias de longitud n. Sea $\oplus$ una operación binaria sobre A tal que para $\mathbf{x}$ y $\mathbf{y}$ en A, $\mathbf{x} \oplus$ y es una sucesión de longitud n que tiene números uno en las posiciones donde $\mathbf{x}$ y $\mathbf{y}$ difieren ~ y números cero en las posiciones donde x y y son iguales. Por ejemplo, sea $\mathbf{x} = 00101$ y $\mathbf{y} = 10110$, entonces $\mathbf{x} \oplus \mathbf{y} = 10011$.

Se deja al lector demostrar que $(A, \oplus)$ es un grupo. Observemos que una palabra con únicamente ceros es la identidad, y que cualquier palabra es su propio inverso en $(A, \oplus)$.

Sea x una palabra en A. Definimos el peso de x, denotado por $w(\mathbf{x})$, como la cantidad de números uno en x. Así, el peso de 1110000 es 3, al igual que el de 1001100. Para x y y en A, definimos la *distancia* entre x y y, denotada por $d(\mathbf{x}, \mathbf{y})$, como el peso de $\mathbf{x} \oplus \mathbf{y}$,

$w(\mathbf{x} \oplus \mathbf{y})$. Por ejemplo, la distancia entre 1110000 y 1001100 es 4, y la distancia entre 1110000 y 0001111 es 7. Observemos que la distancia entre dos palabras es exactamente el número de posiciones en las cuales éstas difieren.

Es obvio que para x y y cualesquiera, $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$. Mostraremos ahora que para Toda $\mathbf{x}, \mathbf{y}, \mathbf{z}$ en A

$$d(\mathbf{x}, \mathbf{y}) \leqq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$$

Claramente, Así,

$$w(\mathbf{u} \oplus \mathbf{v}) \leqq w(\mathbf{u}) + w(\mathbf{v})$$

tenemos que

$$\begin{aligned} w(\mathbf{x} \oplus \mathbf{y}) &= w(\mathbf{x} \oplus \mathbf{z} \oplus \mathbf{z} \oplus \mathbf{y}) \\ &\leqq w(\mathbf{x} \oplus \mathbf{z}) + w(\mathbf{z} \oplus \mathbf{y}) \end{aligned}$$

o

$$d(\mathbf{x}, \mathbf{y}) \leqq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$$

Sea G un código de bloque. Definimos la *distancia* de G como la distancia mínima entre cualquier par de palabras código distintas en G. La distancia de un código de bloque está muy relacionada con su capacidad para corregir errores, como mostraremos más adelante. Supongamos que correspondiendo a la transmisión de una palabra código en G, se ha recibido la palabra y. Nuestro problema es determinar a partir de y la palabra código que fue transmitida. Para motivar nuestro análisis supongamos el caso sencillo en el que y resulta ser una de las palabras código en G. En este caso, uno probablemente se inclinaría por la conclusión obvia de que la palabra transmitida efectivamente fue y. Observemos que esa conclusión obvia aún requiere de justificación. Si suponemos que en el transcurso de la transmisión pueden ocurrir errores en cualquiera de las posiciones, cualquiera de las palabras código en G podría haber sido la palabra transmitida. Cuando supusimos que la palabra transmitida fue y, estábamos suponiendo de modo tácito que cuando la palabra fue transmitida era más probable que no ocurriera error alguno a que hubiesen ocurrido algunos errores. Por ahora acordemos no seguir dichos supuestos y establezcamos una manera general de determinar la palabra transmitida correspondiente a la palabra recibida y. Denotemos por x,, $\mathbf{x}_2, \ldots, \mathbf{x}_N$ a las palabras código en G. Calcularemos la probabilidad condicional $P(\mathbf{x}_i|\mathbf{y})$ para

$i = 1, 2, \ldots, N$, donde $P(\mathbf{x}_i|\mathbf{y})$ es la probabilidad de que $\mathbf{x}_i$ fuese la palabra transmitida ya que y fue la palabra recibida. Si $P(\mathbf{x}_k|\mathbf{y})$ es la mayor de todas las probabilidades condicionales calculadas, concluimos que $\mathbf{x}_k$ fue la palabra transmitida. Este criterio para determinar la palabra transmitida se conoce como *criterio de decodificación de la máxima-probabilidad.*

El cálculo de la probabilidad condicional $P(\mathbf{x}_i|\mathbf{y})$ puede ser realmente complicado ya que la probabilidad depende de muchos factores en el sistema de comunicación. Existe, no obstante, otro criterio que puede utilizarse para determinar la palabra transmitida. Calculamos $d(\mathbf{x}_i, \mathbf{y})$ para $i = 1, 2, \ldots, N$, y concluimos que x_k fue la palabra transmitida si $d(\mathbf{x}_k, \mathbf{y})$ es la menor distancia entre todas las distancias calculadas. Esto se conoce como *criterio de decodificación de la mínima-distancia.* Si suponemos que la presencia de errores en las posiciones son independientes, y que la probabilidad de que haya un error es p, entonces $P(\mathbf{x}_i|\mathbf{y}) = (1-p)^{n-t}p^t$ donde t es la distancia entre $\mathbf{x}_i$ y $\mathbf{y}$. Para $p < \frac{1}{2}$, entre menor sea $d(\mathbf{x}_i, \mathbf{y})$, mayor será $P(\mathbf{x}_i, \mathbf{y})$.† En consecuencia, el criterio de decodificación de la mínima-distancia es equivalente al criterio de decodificación de la máxima-probabilidad (en este caso, la conclusión de que la palabra transmitida fue y cuando la palabra recibida y es una palabra código está justificada).

Señalemos de inmediato que *un código de distancia* $2t + 1$ *puede corregir t o menos errores de transmisión cuando se sigue el criterio de decodificación de la mínima-distancia.* Supongamos que una palabra código x fue transmitida y que la palabra y fue recibida. Si no se han cometido más de t errores en el transcurso de la transmisión, tenemos que

$$d(\mathbf{x}, \mathbf{y}) \leqq t$$

Sea $\mathbf{x}_1$, otra palabra código. Puesto que

$$d(\mathbf{x}, \mathbf{x}_1) \geqq 2t + 1$$

y

$$d(\mathbf{x}, \mathbf{x}_1) \leqq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{x}_1)$$

tenemos que

$$d(\mathbf{y}, \mathbf{x}_1) \geqq t + 1$$

Luego entonces, el criterio de decodificación de la mínima-distancia efectivamente seleccionará a x como la palabra transmitida.

Ahora estudiaremos una clase de códigos de bloque conocida como *códigos de grupo.* Un subconjunto G de A es llamado un código de grupo si $(G, \oplus)$ es un subgrupo de $(A, \oplus)$, donde A es el conjunto de sucesiones binarias de longitud n.

Mostraremos que la distancia de G es igual al peso mínimo de las palabras diferentes del nulo en G. Este resultado hace mucho más simple calcular la distancia de un código de grupo, puesto que ya no es necesario calcular exhaustivamente la distancia para cada par de palabras distintas en G. Supongamos que x es una palabra diferente del nulo en G. Debido a que

$$w(\mathbf{x}) = d(\mathbf{x}, \mathbf{0})^{‡}$$

---

† Véase el problema 11.42 para el caso $p > \frac{1}{2}$.
‡ 0 denota la palabra todo-cero.

y ya que $\mathbf{0}$ está en G, tenemos que

$$w(\mathbf{x}) \geqq \min_{\mathbf{y},\, \mathbf{z} \in G} [d(\mathbf{y}, \mathbf{z})] \tag{11.1}$$

Por otro lado, para cualesquiera y y z en *G,* ya que

$$d(\mathbf{y}, \mathbf{z}) = w(\mathbf{y} \oplus \mathbf{z})$$

$$d(\mathbf{y}, \mathbf{z}) \geqq \min_{\substack{\mathbf{x} \in G \\ \mathbf{x} \neq \mathbf{0}}} [w(\mathbf{x})] \tag{11.2}$$

A partir de (11.1) obtenemos

$$\min_{\substack{\mathbf{x} \in G \\ \mathbf{x} \neq \mathbf{0}}} [w(\mathbf{x})] \geqq \min_{\mathbf{y},\, \mathbf{z} \in G} [d(\mathbf{y}, \mathbf{z})] \tag{11.3}$$

Y a partir de (11.2) obtenemos

$$\min_{\mathbf{y},\, \mathbf{z} \in G} [d(\mathbf{y}, \mathbf{z})] \geqq \min_{\substack{\mathbf{x} \in G \\ \mathbf{x} \neq \mathbf{0}}} [w(\mathbf{x})] \tag{11.4}$$

Al combinar (11.3) y (11.4) llegamos a que

$$\min_{\substack{\mathbf{x} \in G \\ \mathbf{x} \neq \mathbf{0}}} [w(\mathbf{x})] = \min_{\mathbf{y},\, \mathbf{z} \in G} [d(\mathbf{y}, \mathbf{z})]$$

y ya que $\mathbf{y} \oplus \mathbf{z}$ está también en *G,* tenemos que

Para códigos de grupo, existe una manera eficiente para determinar la palabra transmitida correspondiente a una palabra recibida de acuerdo con el criterio de decodificación de la mínima-distancia. Sea $(G, \oplus)$ un código de grupo. Sea y una palabra recibida. Debido a que $d(\mathbf{x}_i, \mathbf{y}) = w(\mathbf{x}_i \oplus \mathbf{y})$, los pesos de las palabras en el conjunto cocientey $G \oplus$ son las distancias entre las palabras código en *G* y y. Denotemos por e a la palabra[†] de menor peso en $G \oplus \mathbf{y}$. Sea $\mathbf{e} = \mathbf{x}_j \oplus \mathbf{y}$ donde $\mathbf{x}_j$ está en *G.* De acuerdo con el criterio de decodificación de la mínima-distancia, $\mathbf{e} \oplus \mathbf{y} = \mathbf{x}_j$ es la palabra código transmitida. Como este argumento es válido para toda y en el conjunto cociente $G \oplus \mathbf{y}$, nuestro procedimiento decodificador puede establecerse como:

1. Determine todos los conjuntos cociente de *G.*
2. Para cada conjunto cociente, tome la palabra de menor peso,[‡] a la cual nos referiremos como el líder del conjunto cociente.
3. Para una palabra recibida $\mathbf{y}$, $\mathbf{e} \oplus$ y es la palabra transmitida, donde e es el líder del conjunto cociente que contiene a y.

Como ejemplo consideremos que *G* = {0000, 0011, 1101, 1110}. Podemos verificar fácilmente que $(G, \oplus)$ es un grupo. Las filas de la figura 11.9 son los distintos conjuntos

† O una de las palabras de menor peso.
‡ O una de las palabras de menor peso.

0000	0011	1101	1110
1000	1011	0101	0110
0100	0111	1001	1010
0010	0001	1111	1100

Figura 11.9

cociente de *G.* Dejamos al lector verificar que de acuerdo con el criterio de decodificación de la mínima-distancia, la palabra recibida 1011 será decodificada como 0011, la palabra recibida 1010 será decodificada como 1110, y la palabra recibida 1111 será decodificada comollOlolllO, según si 0010 o 0001 fuese escogida como el líder del conjunto cociente que contiene a la palabra 1111.

11.8 ISOMORFISMOS Y AUTOMORFISMOS

Sea $(A, \star)$ el sistema algebraico mostrado en la figura 11.10a. Porque tanto los nombres de los elementos de *A, a, b, c* y *d,* como el nombre de la operación sobre A, $\star$ son nombres abstractos, no existe razón alguna por la que no podamos cambiarles por otros nombres abstractos. Por ejemplo, al cambiar a, b, c, d por $\alpha, \beta, \gamma, \delta$ y $\star$ por $*$, obtenemos el sistema Algebraico $(B, *)$ de la figura 11.106. Es claro que uno podría estar de acuerdo con que los dos sistemas $(A, \star)$ y $(B, *)$ son "esencialmente el mismo". Diremos que un sistema algebraico $(B, *)$ es *isomorfo* al sistema algebraico isi$(A, \star)$ podemos obtenera $(B, *)$ partir de $(A, \star)$ mediante el renombrado de los elementos, la operación en $(A, \star)$ o ambas cosas. De manera más formal pero equivalente, decimos que $(B, *)$ es isomorfo asi $(A, \star)$ existe una función biyectiva/de *A* hacia *B* tal que para todo a_1 y a_2 en A

$$f(a_1 \star a_2) = f(a_1) * f(a_2)$$

La función/se conoce como un *isomorfismo* de $(A, \star)$ hacia $(B, *)$, y $(B, *)$ se denomina la *imagen isomorfa* de *A.* Por ejemplo, la función/tal que

$$f(a) = \alpha$$
$$f(b) = \beta$$
$$f(c) = \gamma$$
$$f(d) = \delta$$

$\star$	a	b	c	d
a	a	b	c	d
b	b	a	a	c
c	b	d	d	c
d	a	b	c	d

a)

$*$	α	β	γ	δ
α	α	β	γ	δ
β	β	α	α	γ
γ	β	δ	δ	γ
δ	α	β	γ	δ

b)

Figura 11.10

$\star$	a	b
a	a	b
b	b	a

$(A, \star)$

$\oplus$	PAR	IMPAR
PAR	PAR	IMPAR
IMPAR	IMPAR	PAR

$(B, \oplus)$

$*$	0°	180°
0°	0°	180°
180°	180°	0°

$(C, *)$

+	0¢	5¢
0¢	0¢	5¢
5¢	5¢	0¢

$(D, +)$

Figura 11.11

es un isomorñsmo del sistema algebraico $(A, \star)$ en la figura 11.10a hacia el sistema algebraico $(B, *)$ en la figura 11.10¿>. Observemos que la función *g* tal que

$$g(a) = \delta$$
$$g(b) = \gamma$$
$$g(c) = \beta$$
$$g(d) = \alpha$$

también es un isomorfismo desde $(A, \star)$ hacia $(B, *)$.

Otro ejemplo, en la figura 11.11 los sistemas algebraicos $(B, \oplus)$, $(C, *)$ y $(D, +)$ son todos isomorfos al sistema algebraico $(A, \star)$.De hecho, el sistemacorresponde$(B, \oplus)$ a la adición de números pares e impares, el sistema$(C, *)$corresponde a la rotación de figuras geométricas en el plano por 0° y 180°, y el sistema *(D,* +) a la situación de comprar dos productos en una tienda de cinco y diez centavos, donde el elemento 0¢ en D se usa para cantidades que son múltiplos de 10¢, el elemento 5¢ se usa para cantidades que son múltiplos de 10¢ más 5¢, y la operación binaria + determina si el precio total de la compra es un múltiplo de 10¢ o un múltiplo de 10¢ más 5¢. En realidad, la noción de isomorfismo entre
dos sistemas algebraicos relaciona los sistemas algebraicos abstractos a situaciones físicas que encontramos en la práctica. En consecuencia, las propiedades de los sistemas algebraicos abstractos tendrán una interpretación directa en términos de situaciones físicas. Desde el punto de vista de las aplicaciones, ésta es en realidad *la* razón para estudiar sistemas algebraicos abstractos.

Un isomorfismo de un sistema algebraico $(A, \star)$ hacia $(A, \star)$ se denomina *automorfismo* sobre $(A, \star)$.Por ejemplo, la función/tal que

$$f(a) = d$$
$$f(b) = c$$
$$f(c) = b$$
$$f(d) = a$$

es un automorfísmo sobre el sistema algebraico $(A, \star)$ en la figura 11.1 *Oa*. Una interpretación física de un automorfísmo sobre un sistema algebraico es una manera en la cual los elementos en el sistema intercambian sus papeles.

Ejemplo 11.4

Como ejemplo ilustrativo sobre la noción de isomorfismo entre grupos, mostraremos que existe un grupo único, salvo un isomorfismo, de orden *p* para cualquier primo *p*. Recordemos que para cualquier entero n, $(Z_n, \oplus)$ es un grupo. Sea $(G, \star)$ un grupo de orden *p*. Puesto que cualquier grupo de orden primo es cíclico, los elementos en *G* pueden representarse como $a^0, a, a^2, \ldots, a^{p-1}$ para cualquier elemento *a* en G que no sea la identidad.[1] La función/(a') = *i* es claramente un isomorfismo desde $(G, \star)$ hacia $(Z_p, \oplus)$. Por tanto, concluimos que cualquier grupo de orden *p* es isomorfo a $(Z_p, \oplus)$. ❐

11.9 HOMOMORFISMOS Y SUBGRUPOS NORMALES

La noción de sistemas algebraicos isomorfos puede generalizarse fácilmente. Sean $(A, \star)$ y $(B, *)$ dos sistemas algebraicos. Sea/una función de *A* hacia *B* tal que para cualesquiera a, y a_2 en *A*

$$f(a_1 \star a_2) = f(a_1) * f(a_2)$$

/se denomina un *homomorfismo* de $(A, \star)$ hacia $(B, *)$, y $(B, *)$ se denomina la *imagen homomorfa* de $(A, \star)$.

Por ejemplo, para los dos sistemas algebraicos de las figuras 11.12o y *b*, la función/ tal que

$$f(\alpha) = 1 \qquad f(\beta) = 1 \qquad f(\gamma) = 1$$
$$f(\delta) = 0 \qquad f(\varepsilon) = 0$$
$$f(\zeta) = -1$$

$\star$	α	β	γ	δ	ε	ζ
α	α	β	α	α	γ	δ
β	β	α	γ	β	γ	ε
γ	α	γ	α	β	γ	ε
δ	α	β	β	δ	ε	ζ
ε	γ	γ	γ	ε	ε	ζ
ζ	δ	ε	ε	ζ	ζ	ζ

a)

$*$	1	0	−1
1	1	1	0
0	1	0	−1
−1	0	−1	−1

b)

Figura 11.12

† a^0 representa la identidad de $(G, \star)$.

★	a	b	c	d
a	a	a	d	c
b	b	a	c	d
c	c	d	a	b
d	d	d	b	a

a)

	a	b	c	d
a	√	√		
b	√	√		
c			√	√
d			√	√

b)

Figura 11.13

es un homomorfismo del sistema algebraico $(\{\alpha, \beta, \gamma, \delta, \varepsilon, \zeta\}, \star)$ hacia el sistema algebraico $(\{1, 0, -1\}, *)$.

Existe otra manera de revisar la noción de un homomorfismo de un sistema algebraico hacia otro. Consideremos el sistema algebraico $(A, \star)$ y una relación de equivalencia R sobre A. Denominamos a R una *relación de congruencia* sobre A (respecto a $\star$) si (a_1, a_2) y (b_1, b_2) en R implican que $(a_1 \star b_1, a_2 \star b_2)$ también esté en R. Por ejemplo, para el sistema algebraico de la figura 11.13a, la relación de equivalencia R en la figura 11.136 es una relación de congruencia. Por otro lado, para el sistema algebraico de la figura 11.14a, la relación de equivalencia R de la figura 11.146 no es una relación de congruencia [ya que, aun cuando (a, b) y (c, d) están en R, $(a \star c, b \star d)$ lo cual es igual a (d, a), no está en R].

Las clases de equivalencia en las que se divide A son llamadas *clases de congruencia.* Podemos definir un nuevo sistema algebraico $(B, *)$ de la siguiente manera: sea $B = \{A_1, A_2, \ldots, A_r\}$ la partición de A inducida por R.† Sea $*$ la operación binaria tal que para A_i y A_j cualesquiera en B, $A_i * A_j$ es igual a la clase de congruencia A_k que contiene al elemento $a_1 \star a_2$, donde a_1 es cualquier elemento en A_i y a_2 es cualquier elemento en A_j. Observemos que debido a que R es una relación de congruencia, la operación $*$ está bien definida (esto es, la clase de congruencia $A_i * A_j$ está determinada de manera única sin importar qué elementos a_1 en A_i y a_2 en A_j se hayan tomado). $(B, *)$ es una imagen homomorfa de $(A, \star)$, ya que la función

$$f(a) = A_i \qquad \text{si} \qquad a \in A_i$$

es un homomorfismo desde $(A, \star)$ hacia $(B, *)$.‡ Por intuición, una imagen homomorfa de un sistema algebraico puede verse como una descripción global del comportamiento del sistema cuando se ignoran algunas de las características que distinguen a ciertos elementos

★	a	b	c	d
a	a	a	d	c
b	b	a	d	a
c	c	b	a	b
d	d	d	b	a

	a	b	c	d
a	√	√		
b	√	√		
c			√	√
d			√	√

Figura 11.14

† Recordamos al lector que una partición es un conjunto de subconjuntos y que es perfectamente correcto que los elementos del conjunto B en el sistema algebraico $(B, *)$ sean a su vez conjuntos.

‡ Exhortamos al lector a mostrar la recíproca, a saber, un homomorfismo de $(A, \star)$ hacia $(B, *)$ induce una relación de congruencia sobre A.

$\odot$	POSITIVO	NEGATIVO	CERO
POSITIVO	POSITIVO	NEGATIVO	CERO
NEGATIVO	NEGATIVO	POSITIVO	CERO
CERO	CERO	CERO	CERO

Figura 11.15

en el sistema. En consecuencia, los elementos que se vuelven indistinguibles pueden ser colocados en una clase de congruencia, y el comportamiento del sistema puede describirse por la manera en que interactúan estas clases de congruencia.

Como ejemplo consideremos el sistema algebraico de la figura 11.12a como una descripción de la interacción de seis tipos diferentes de partículas $\{\alpha, \beta, \gamma, \delta, \varepsilon, \zeta\}$. Supongamos que α, β y γ son partículas de carga positiva; δ y ε son partículas neutras; y ζ es una partícula de carga negativa. Si $\{1, 0, -1\}$ denota los tres tipos de partícula, la figura 11. \2*b* muestra cómo interactúan los tres tipos de partículas. En efecto, como se señaló con anterioridad, la función/tal que

$$\begin{array}{lll} f(\alpha) = 1 & f(\beta) = 1 & f(\gamma) = 1 \\ f(\delta) = 0 & f(\varepsilon) = 0 & \\ f(\zeta) = -1 & & \end{array}$$

es un homomorfísmo del sistema algebraico $(\{\alpha, \beta, \gamma, \delta, \varepsilon, \zeta\}, \star)$ hacia el sistema algebraico $(\{1, 0, -1\}, *)$.

Otro ejemplo, consideremos el sistema algebraico (/, •), donde /es el conjunto de todos los enteros y • es la operación ordinaria de multiplicación de enteros. Supongamos que no estamos interesados en distinguir a todos los enteros en *I*, sino que deseamos hacer una distinción entre enteros positivos, enteros negativos y cero. De inmediato observamos que el sistema algebraico $(B, \odot)$ en la figura 11.15 es una imagen homomorfa de (/, •), y que la función/tal que

$$f(n) = \begin{cases} \text{POSITIVO} & \text{si } n \text{ es un entero positivo} \\ \text{NEGATIVO} & \text{si } n \text{ es un entero negativo} \\ \text{CERO} & \text{si } n = 0 \end{cases}$$

es efectivamente un homomorfísmo de $(I, \cdot)$ **hacia** $(B, \odot)$.

Dado un grupo $(G, \star)$ es posible saber cómo determinar una imagen homomorfa de $(G, \star)$ y más aún, todas las imágenes homomorfas de $(G, \star)$.† Recordemos de nuestro análisis de la sección 11.5 que un subgrupo de $(G, \star)$ induce una partición de *G* en los conjuntos cocientes del subgrupo. Entonces resulta razonable preguntarse si en efecto esta partición divide *G* en clases de congruencia. La respuesta a esta pregunta es negativa, aunque una condición adicional sobre el subgrupo haría el truco.

---

† No es difícil ver que la imagen homomorfa de un grupo siempre es un grupo. Véase el problema 11.36.

★	α	β	γ	δ	ε	ζ
α	α	β	γ	δ	ε	ζ
β	β	γ	α	ε	ζ	δ
γ	γ	α	β	ζ	δ	ε
δ	δ	ζ	ε	α	γ	β
ε	ε	δ	ζ	β	α	γ
ζ	ζ	ε	δ	γ	β	α

$(A, \star)$

a)

*	{α, β, γ}	{δ, ε, ζ}
{α, β, γ}	{α, β, γ}	{δ, ε, ζ}
{δ, ε, ζ}	{δ, ε, ζ}	{α, β, γ}

$(B, *)$

b)

Figura 11.16

Sea//un subgrupodeG. Diremos que *H* es un *subgrupo normal* si, para todo elemento *a* en *G,* el conjunto cociente izquierdo $a \star H$ es igual al conjunto cociente derecho $a\ H \star$ (si *G* es un grupo conmutativo, cualquier subgrupo de *G* es normal). Por ejemplo, para el grupo $(A, \star)$ mostrado en la figura 11.16a, $(H, \star)$ es un subgrupo normal, donde $H = \{\alpha, \beta, \gamma\}$. Por ejemplo,

$$\delta \star H = \delta \star \alpha, \delta \star \beta, \delta \star \gamma\} = \{\delta, \zeta, \varepsilon\}$$

$$H \star \delta = \{\alpha \star \delta, \beta \star \delta, \gamma \star \delta\} = \{\delta, \varepsilon, \zeta\}$$

A continuación queremos demostrar que los distintos conjuntos cociente izquierdos (de rechos) de un subgrupo normal *H* son clases de congruencia de *G*. Sea $a \star H$ y $b \star H$_ dos conjuntos cociente. Queremos demostrar que para todos los elementos a_1 en $a \star H$ y todos los
elementos b_1 en $b \star H$, los elementos $a_1 \star b_1$ están todos en un conjunto cociente de *H*. Sean

$$a_1 = a \star h_1$$

$$b_1 = b \star h_2$$

para algún h_1 y h_2 en H. Tenemos que

a_x • b_l = (a • /),) • (b • h_2) = (a •
h_x) • (/z_3 * b) - a*   para algún /?$_3$ 6 *H*
h_4* b = a-k b-k h_5 para algún h_4 e *H*
para algún h_5 e *H*

Así, a, • b_x, está en el conjunto cociente (a • *b)* • *H.*

Por ejemplo, para el grupo *{A,* •) mostrado en la figura 11.16a y el subgrupo normal (//, •), donde *H* = {a, p, y}, las clases de congruencia son {a, p, y} y {5, s, *Q*. En consecuencia, tenemos la imagen homomorfa *(B,* *) mostrada en la figura 11.166, donde *B*= {{a, p, y},{5, E, *Q}*. Como otro ejemplo, para el grupo (/, +), donde /es el conjunto de todos los enteros y + es la operación ordinaria de adición de enteros, *(E, +)* es un subgrupo

$\oplus$	PAR	IMPAR
PAR	PAR	IMPAR
IMPAR	IMPAR	PAR

Figura 11.17

normal donde E es el conjunto de todos los enteros pares. Correspondiendo al subgrupo $(E, +)$, tenemos la imagen homomorfa mostrada en la figura 11.17, donde PAR es clase de congruencia de todos los enteros pares, e IMPAR es la clase de congruencia de todos los enteros impares.

Nuestro siguiente paso es determinar si al agotar todos los subgrupos normales de$(A, \star)$ se originará el agotamiento de todas las imágenes homomorfas de$(A, \star)$.Resulta que efectivamente esto es lo que sucede, como mostraremos adelante. Sea/un homomorfismo de$(A, \star)$ hacia $(B, *)$. Queremos demostrar que/corresponde a una partición de A en clases de congruencia inducidas por un subgrupo normal. Sea //el conjunto de todos los elementos en A cuyas imágenes bajo/son la identidad de B, a la cual denotaremos por s.

Primero demostramos que$(H, \star)$ es un subgrupo de $(A, \star)$:

1. Queremos demostrar que • es cerrada sobre //. Para todo a, b en H, ya que

$$f(a \star b) = f(a) * f(b) = \varepsilon * \varepsilon = \varepsilon$$

$a \star b$ está también en //.

2. Queremos demostrar que e, la identidad de A, está en //. Para un elemento arbitrario a en A tenemos que

$$f(a \star e) = f(a) * f(e)$$

o

$$f(a) = f(a) * f(e)$$

Puesto que $(B, *)$es un grupo, $f(e)$ debe ser la identidad de $(B, *)$. Por tanto, e está en//.

3. Queremos demostrar que o"[1] está en //para cualquier elemento a en H. Puesto que

$$f(a \star a^{-1}) = f(a) * f(a^{-1})$$

o

$$f(e) = f(a) * f(a^{-1})$$

$$\varepsilon = \varepsilon * f(a^{-1})$$

o

$$\varepsilon = f(a^{-1})$$

o concluimos que a ' está en H.

Así, $a \star h \star a^{-1}$ está en H. Esto es, $a \star h \star a^{-1} = h_1$ o $a \star h = h_1 \star a$ para algún h_x en H. De

Ahora demostraremos que (//, •) es un subgrupo normal: para cualquier *aenAyh* en//

$$
\begin{aligned}
f(a \star h \star a^{-1}) &= f(a) * f(h) * f(a^{-1}) \\
&= f(a) * \varepsilon * f(a^{-1}) \\
&= f(a) * f(a^{-1}) \\
&= f(a \star a^{-1}) \\
&= f(e) \\
&= \varepsilon
\end{aligned}
$$

donde concluimos que //es normal.

Finalmente, demostraremos que si a y b están en el mismo conjunto cociente de H, entonces $f(a) = f(b)$. Puesto que b puede escribirse como

$$b = a \star h$$

para algún h en H, tenemos que

$$f(b) = f(a) * f(h) = f(a)$$

Recíprocamente, queremos demostrar que, si/(a) =*f(b),* entonces a y b están en el mismo conjunto cociente de H. Puesto que

$$
\begin{aligned}
f(a^{-1} \star b) &= f(a^{-1}) * f(b) \\
&= f(a^{-1}) * f(a) \\
&= \varepsilon
\end{aligned}
$$

$a^{-1} \star b$ está en H, esto es, $a^{-1} \star b = h$ o $b = a \star h$ para algún h en H.

11.10 ANILLOS, DOMINIOS INTEGRALES Y CAMPOS

Hasta el momento hemos estudiado sistemas algebraicos con una sola operación binaria. Ahora continuaremos con un breve estudio de varias clases de sistemas algebraicos con dos operaciones binarias. Es claro que dados dos sistemas algebraicos $(A, \star)$ y $(A, *)$, siempre podemos "combinarlos" para dar lugar a un sistema algebraico con dos operaciones binarias $(A, \star, *)$. Por otro lado, un sistema algebraico con dos operaciones puede ser un simple "conglomerado" de dos sistemas algebraicos con una operación si las dos operaciones están relacionadas de alguna manera. Por ejemplo, sea $(A, \star, *)$ un sistema algebraico tal que para

a y b en *A,* cualesquiera, $a \star b = b * a$. Observemos que las dos operaciones $\star$ y $*$ están relacionadas en que ambas son conmutativas, o bien ambas no lo son (¿por qué?). Una manera más natural e importante en la cual dos operaciones binarias pueden estar relacionadas es la propiedad de *distributividad.* Sea $(A, \star, *)$ un sistema algebraico con dos

★	α	β
α	α	β
β	β	α

*	α	β
α	α	α
β	α	β

Figura 11.18

operaciones binarias. Se dice que la operación * es *distributiva* sobre la operación ★ si para todo *a, b y c* en *A,*

$$a * (b \star c) = (a * b) \star (a * c)$$

y

$$(b \star c) * a = (b * a) \star (c * a)$$

Por ejemplo, para el sistema algebraico $(\{\alpha, \beta\}, \star, *)$ donde ★ y * están definidas en la figura 11.18, la operación * es distributiva sobre la operación ★, en tanto que la operación • *no* es distributiva sobre la operación.†

Estamos particularmente interesados en las clases de sistemas algebraicos con dos operaciones los cuales son conocidos como *anillos, dominios integrales* y *campos.* Un sistema algebraico $(A, +, \cdot)$ se denomina un *anillo* si satisface las siguientes condiciones:

1. *{A,* +) es un grupo abeliano.
2. *{A,* •) es un semigrupo.
3. La operación ■ es distributiva sobre la operación +.

Sea Z,, el conjunto de los enteros $\{0, 1, 2, \ldots, n-1\}$. Sea $\oplus$ una operación binaria sobre Z_n, tal que

$$a \oplus b = \begin{cases} a+b & \text{si} \quad a+b < n \\ a+b-n & \text{si} \quad a+b \geqq n \end{cases}$$

Sea © una operación binaria sobre Z_n tal que

$$a \odot b = \text{residuo de dividir } ab \text{ por } n$$

Como se señaló en la sección 11.12, $(Z_n, \oplus)$ es un grupo abeliano. Además, no es difícil ver que $\odot$ es una operación cerrada y asociativa.‡ En consecuencia, $(Z_n, \odot)$ es un semigrupo. El lector demostrará que $\odot$ se distribuye sobre ® y así concluirá que $(Z_n, \oplus, \odot)$ es un anillo.

De modo consistente nos referiremos a las dos operaciones de un anillo $(A, +, \cdot)$ como a las operaciones de adición y multiplicación. También nos referimos *aa* + *b* como a la *suma* de *a* y *b,* y a *a . b* como al *producto* de *a* y *b.* La identidad del grupo abeliano *{A,* +) es conocida como *identidad aditiva* y se denota por 0. El inverso de un elemento *a* del grupo *{A,* +) se denomina como el *inverso aditivo* de a y se denota por *-a.* Algunas veces escribimos *a* - *b* para indicar la suma de a y el inverso aditivo de *b.*

† Por ejemplo, $\beta \star (\alpha * \beta) = \beta$ y $(\beta \star \alpha) * (\beta \star \beta) = \alpha$.
‡ Véase el problema 11.4.

Mostraremos ahora que para todo elemento a en el anillo $(A, +, \cdot)$, $0 \cdot a = a \cdot 0 = 0$. Puesto que

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$$

tenemos que

$$0 \cdot a = 0$$

Que $a \cdot 0 = 0$ puede demostrarse de manera similar.†

Sea $(A, +, \cdot)$ un sistema algebraico con dos operaciones binarias. Se dice que $(A, +, \cdot)$ es un *dominio integral* si:

1. $(A, +)$ es un grupo abeliano.
2. La operación es conmutativa. Y además, si $c \neq 0$ y $c \cdot a = c \cdot b$, entonces $a = b$, donde 0 denota la identidad aditiva.
3. La operación • es distributiva sobre la operación +.

Por ejemplo, sean A el conjunto de todos los enteros, y + y • son las operaciones ordinarias de adición y multiplicación de enteros. Observemos que $(A, +)$ es un grupo abeliano con 0 como la identidad aditiva y $-n$ como el inverso aditivo de n para cualquier entero n; también notamos que la operación • es conmutativa. Además, para cualquier entero diferente de cero c, $c \cdot a = c \cdot b$ implica que $a = b$. Puesto que • es distributiva sobre +, se tiene que $(A, +, \cdot)$ es un dominio integral.

Sea $(A, +, \cdot)$ un sistema algebraico con dos operaciones binarias. Se dice que $(A, +, \cdot)$ es un *campo* si:

1. $(A, +)$ es un grupo abeliano.
2. $(A - \{0\}, \cdot)$ es un grupo abeliano.
3. La operación • es distributiva sobre la operación +.

Exhortamos al lector a verificar que $(F, +, \cdot)$ es un campo, donde F es el conjunto de todos los números racionales, y + y • son las operaciones ordinarias de adición y multiplicación de números racionales. Otros ejemplos de campos son $(R, +, \cdot)$, donde R es el conjunto de todos los números reales y + y • son las operaciones ordinarias de adición y multiplicación de números reales, y $(C, +, \cdot)$, donde C es el conjunto de todos los números com y + y • son las operaciones ordinarias de adición y multiplicación de números complejos (es probable que se dé cuenta en este momento, si no ha sucedido antes, de que cuando estudiamos la adición, sustracción, multiplicación y división de números racionales, reales y complejos en la educación primaria y secundaria, estudiamos las operaciones en los campos de números racionales, reales y complejos. Además, en realidad la sustracción no es una operación "independiente" ya que equivale a la adición del inverso aditivo de un elemento. De modo similar, la división es equivalente a la multiplicación por el inverso multiplicativo de un elemento).

† Observe, en efecto, cuan "útil" es la propiedad distributiva. Véase también el teorema 12.9.

Mostraremos otro ejemplo en que el sistema algebraico $(Z_n, \oplus, \odot)$ es un campo si y sólo si *n* es un número primo. Es claro que tanto $\oplus$ como $\odot$ son operaciones conmutativas. Recordemos que $(Z_n, \oplus)$ es un grupo para todo *n.* Ahora mostraremos quees $(Z_n - \{0\}, \odot)$ un grupo si y sólo si *n* es un número primo. Si *n* no es un número primo, *n* = *ab* para algún a y b en $Z_n - \{0\}$. Esto es, $a \odot b = 0$. Puesto que la operación $\odot$ no es cerrada sobre $Z_n - \{0\}$, $(Z_n - \{0\}, \odot)$ no puede ser un grupo. Examinemos el caso cuando *n* es un número primo. Primero, para todo a y b en $Z_n - \{0\}$, $a \odot b$ no es igual a 0. Por consiguiente, la operación $\odot$ es cerrada sobre el conjunto $Z_n - \{0\}$. Además, es claro que la operación $\odot$ es una operación asociativa y que 1 es la identidad de $(Z_n - \{0\}, \odot)$. Finalmente, mostraremos que para cualquier *a* y dos elementos distintos cualesquiera b y c en $Z_n - \{0\}$, $a \odot b$ no es igual a $a \odot c$. Supongamos que $a \odot b$ es igual a $a \odot c$. Esto es,

$$ab = kn + r$$
$$ac = ln + r$$

para algunos *k,lyr*. Sin pérdida de generalidad, supongamos que $b > c$ y $k > l$. Tenemos que

$$ab - ac = kn - ln$$

o

$$a(b - c) = (k - l)n$$

Puesto que tanto *a* como *b* - *c* son menores que «,(11.5) no es posible cuando *n* es primo. Ahora se tiene del principio del palomar que, para cualquier *a* en $Z_n - \{0\}$, existe un *b* en $Z_n - \{0\}$ tal que

$$a \odot b = 1$$

Esto es, *b* es el inverso de *a* en $(Z_n - \{0\}, \odot)$. Así concluimos qui $(Z_n - \{0\}, \odot)$ es un grupo abeliano. Dejamos al lector demostrar que la operación $\odot$ es distributiva sobre la operación $\oplus$.

Para un número primo *n,* usualmente se conoce $(Z_n, \oplus, \odot)$ como el *campo de enteros módulo n.*

La identidad del grupo $(A - \{0\}, \cdot)$ se conoce como la *identidad multiplicativa,* la cual se denota por 1. El *inverso* de un elemento *a* en un grupo $(A - \{0\}, \cdot)$ se conoce como el *inverso multiplicativo* de a y se denota por $1/a$. Algunas veces escribimos *alb* para indicar el producto de a y el inverso multiplicativo de *b.*

•11.11 HOMOMORFISMOS DE ANILLOS

En la sección 11.9 presentamos la noción de homomorfismo sobre sistemas algebraicos con una operación binaria. Ahora estudiaremos homomorfismos sobre sistemas algebraicos con dos operaciones y, en particular, con mayor detalle, sobre anillos. Sean $(A, +, \cdot)$ y $(B, \oplus, \odot)$

$\oplus$	PAR	IMPAR
PAR	PAR	IMPAR
IMPAR	IMPAR	PAR

$\odot$	PAR	IMPAR
PAR	PAR	PAR
IMPAR	PAR	IMPAR

Figura 11.19

dos sistemas algebraicos. Una función / sobreyectiva, de A hacia B se dice que es un *homomorfismo de* $(A, +, \cdot)$ *hacia* si $(B, \oplus, \odot)$ para cualesquiera a y b en A

$$\begin{aligned} f(a+b) &= f(a) \oplus f(b) \\ f(a \cdot b) &= f(a) \odot f(b) \end{aligned}$$

Además, $(B, \oplus, \odot)$ es denominado como una *imagen homomorfa* de $(A, +, \cdot)$.

En modo alternativo, sea R una relación de congruencia sobre A respecto a ambas operaciones $+$ y $\cdot$. Esto es, R es una relación de equivalencia sobre A y si (a_1, a_2) y (b_1, b_2) están en R, entonces $(a_1 + b_1, a_2 + b_2)$ y $(a_1 \cdot b_1, a_2 \cdot b_2)$ también están en R. Sea $B = \{A_1, A_2, \ldots, A_r\}$ el conjunto de clases de congruencia en las cuales se divide A. Definimos dos operaciones binarias $\oplus$ y $\odot$ sobre B de manera que $A_i \oplus A_j$ es igual a la clase de congruencia donde está $a_1 + a_2$, y $A_i \odot A_j$ es igual a la clase de congruencia donde está $a_1 \cdot a_2$, para todo a_1 en A_i y a_2 en A_j. De acuerdo con nuestro análisis de la sección 11.9, tenemos qué $(B, \oplus, \odot)$ es una imagen homomorfa de $(A, +, \cdot)$.

Por ejemplo, sea N el conjunto de todos lo números naturales, y $+$ y $\cdot$ las operaciones ordinarias de adición y multiplicación de números naturales. Consideremos el sistema algebraico $(\{\text{PAR}, \text{IMPAR}\}, \oplus, \odot)$, donde las operaciones $\oplus$ y $\odot$ están definidas en la figura 11.19. Señalemos que $(\{\text{PAR}, \text{IMPAR}\}, \oplus, \odot)$ es una imagen homomorfa de $(\mathrm{N}, +, \cdot)$. Además, la función/tal que

$$f(n) = \begin{cases} \text{PAR} & \text{si } n \text{ es un número par} \\ \text{IMPAR} & \text{si } n \text{ es un número impar} \end{cases}$$

es el correspondiente homomorfismo de $(\mathrm{N}, +, \cdot)$ hacia $(\{\text{PAR}, \text{IMPAR}\}, \oplus, \odot)$.

Ahora investiguemos el problema de la determinación de las imágenes homomorfas de un anillo.^ Sea $(A, +, \cdot)$ un anillo y $(H, +)$ un subgrupo de $(A, +)$. Sea R la relación de equivalencia que induce H sobre A. Debido a que $(A, +)$ es un grupo abeliano, R es una relación de congruencia respecto a $+$. Queremos conocer bajo qué condiciones R es también una relación de congruencia respecto de •. Supongamos que (a, b) y (c, d) están en R. Esto es

$$\begin{aligned} a &= b + h_1 \\ c &= d + h_2 \end{aligned}$$

para algunos h_x y h_2 en H. De esto se tiene que

$$\begin{aligned} a \cdot c &= (b + h_1) \cdot (d + h_2) \\ &= (b \cdot d) + (b \cdot h_2) + (h_1 \cdot d) + (h_1 \cdot h_2) \end{aligned}$$

t No es difícil demostrar que la imagen homomorfa de un anillo siempre es un anillo (véase el problema 11.36).

Observemos que, *si* tanto *b* ■ h_2 como *h*^ ■ *d* están en *H*, entonces

$$(b \cdot h_2) + (h_1 \cdot d) + (h_1 \cdot h_2)$$

está también en *H*, y tendremos que $(a \cdot c, b \cdot d)$ está en *R*.

Dicha observación motiva la definición de un ideal en un anillo. Sea i$(A, +, \cdot)$ un anillo y *H* un subconjunto de *A*. Diremos que *H* es un *ideal* si se satisfacen las siguientes condiciones:

1. $(H, +)$ es un subgrupo de $(A, +)$.
2. Para todo a en A y h en H, $a \cdot h$ y $h \cdot a$ están en H.†

Se tiene de inmediato, a partir de nuestro análisis, que un ideal de un anillo $(A, +, \cdot)$ induce una relación de congruencia sobre *A* respecto de ambas operaciones $+$ y $\cdot$, y en consecuencia define una imagen homomorfa de $(A, +, \cdot)$.

Por otro lado, demostraremos que cualquier homomorfismo de un anillo corresponde a un ideal del anillo. Sea/un homomorfismo de un anillo $(A, +, \cdot)$ hacia $(B, \oplus, \odot)$. Sea *H* el conjunto de elementos de *A* que son mapeados hacia la identidad aditiva de $(B, \oplus, \odot)$ bajo /[recordamos al lector que como se trata de una imagen homomorfa de $(A, +, \cdot)$, $(B, \oplus, \odot)$ es un anillo]. Ya se ha demostrado en la sección 11.9 que *H* es un subgrupo normal de $(A, +)$. Para demostrar que *H* es un ideal, señalemos que para todo *aenAyhenH,*

$$f(a \cdot h) = f(a) \odot f(h) = f(a) \odot 0 = 0$$

donde 0 denota la identidad aditiva de $(B, \oplus, \odot)$., Así, *a* ■ *h* está en *H*. De manera similar, podemos demostrar que *h* ■ *a* está en *H*. Así, concluimos que *H* es un ideal.

$\oplus$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\odot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

a)

$\boxplus$	{0, 2, 4}	{1, 3, 5}
{0, 2, 4}	{0, 2, 4}	{1, 3, 5}
{1, 3, 5}	{1, 3, 5}	{0, 2, 4}

$\boxdot$	{0, 2, 4}	{1, 3, 5}
{0, 2, 4}	{0, 2, 4}	{0, 2, 4}
{1, 3, 5}	{0, 2, 4}	{1, 3, 5}

b)

Figura 11.20

† Por intuición afirmamos que cualquier elemento en *A* es "deglutido" del ideal cuando es multiplicado por un elemento del ideal.

Por ejemplo, consideremos el anillo$(Z_n, \oplus, \odot)$i mostrado en la figura 11.20a. Podemos verificar de inmediato que {0,2,4} es un ideal. Las dos clases de congruencia son {0,2,4} y {1,3,5}, y la imagen homomorfa se muestra en la figura 11.206.

Consideremos otro ejemplo, el anillo $(I, +, \cdot)$,donde/es el conjunto de todos los enteros, y + y • son las operaciones ordinarias de adición y multiplicación de enteros. Sea *H* el conjunto de todos los múltiplos de$n$ $\{\ldots, -2n, -n, 0, n, 2n, 3n, \ldots\}$ para cualquier entero *n*. Es obvio que //es un ideal. Además, una imagen homomorfa correspondiente es$(Z_n, \oplus, \odot)$.

*11.12 ANILLOS POLINOMIALES Y CÓDIGOS CÍCLICOS

Sea$(A, \star, *)$un sistema algebraico donde • es una operación *m-aria y* * es una operación «-aria. Algunas veces nos referiremos a los elementos de *A* como *constantes*. Una *variable* (o un *indeterminado)* es un nombre simbólico que no es el nombre de algún elemento de *A*. Definimos las *expresiones algebraicas* en$(A, \star, *)$de acuerdo con las siguientes reglas:

1. Cualquier constante es una expresión algebraica.
2. Cualquier variable es una expresión algebraica.
3. Si $e_1, e_2, \ldots, e_m$ son expresiones algebraicas, entonces $\star(e_1, e_2, \ldots, e_m)$ también lo es.
4. Si $e_1, e_2, \ldots, e_n$ son expresiones algebraicas, entonces $*(e_1, e_2, \ldots, e_n)$ también lo es.

Por ejemplo, si *x* y *y* son variables, las siguientes expresiones son expresiones algebraicas en$(Z_4, \oplus, \odot)$:

$$2$$

$$2 \oplus x$$

$$(3 \odot (x \oplus x)) \odot (x \odot y)$$

Sea *{A,* +, ■) un anillo. Una expresión algebraica en *(A,* +, •) de la forma

$$a_0 + (a_1 \cdot x) + (a_2 \cdot x^2) + \cdots + (a_{n-1} \cdot x^{n-1}) + (a_n \cdot x^n)^\dagger \qquad n \geqq 0 \tag{11.6}$$

se conoce como un *polinomio*, donde$a_0, a_1, \ldots, a_{n-1}, a_n$ son constantes y *x* es una variable. Debido a que + y $\cdot$ son operaciones asociativas, la expresión en (11.6) no es ambigua. Usaremos la notación de la forma$A(x), B(x), \ldots$para polinomios. El *grado* de un polinomio

† Usamos x^i para denotar $\underbrace{x \cdot x \cdot \cdots x}_{i \text{ veces}}$.

es el mayor n para el cual a_n es diferente de 0. A partir de la convención de omitir el símbolo de la operación de multiplicación y los paréntesis en (11.6), reescribimos (11.6) como

$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + a_nx^n$$

Analizaremos brevemente la construcción de anillos polinomiales. Sea $(F, +, \cdot)$ un campo. Denotamos por $F[x]$ al conjunto de todos los polinomios en $(F, +, \cdot)$ cuya variable es x. Definimos un sistema algebraico $(F[x], \boxplus, \boxdot)$ donde $\boxplus$ y $\boxdot$ son dos operaciones binarias definidas como sigue: para

$$A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$
$$B(x) = b_0 + b_1x + b_2x^2 + \cdots + b_mx^m$$

tenemos que

$$A(x) \boxplus B(x) = C(x) = c_0 + c_1x + c_2x^2 + \cdots + c_rx^r$$

donde

$$c_i = a_i + b_i^\dagger$$

y

$$A(x) \boxdot B(x) = D(x) = d_0 + d_1x + d_2x^2 + \cdots + d_rx^r$$

donde

$$d_i = (a_0 \cdot b_i) + (a_i \cdot b_{i-1}) + \cdots + (a_i \cdot b_0)$$

Es fácil verificar que $(F[x], \boxplus, \boxdot)$ es un anillo, el cual se conoce como un *anillopolinomial.*

Sea $(F, +, \cdot)$ un campo. Denotemos por $F_n[x]$ al conjunto de polinomios de grado menor que n en $F[x]$. Sea $G(x)$ un polinomio de gradoun n en $F[x]$. Sea $(F_n[x], \blacktriangle, \triangle)$ sistema algebraico, donde A y A son dos operaciones binarias definidas como sigue: para

$$A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$
$$B(x) = b_0 + b_1x + b_2x^2 + \cdots + b_{n-1}x^{n-1}$$

tenemos que

$$A(x) \blacktriangle B(x) = A(x) \boxplus B(x)$$

y

$$A(x) \triangle B(x) = \text{el residuo de } A(x) \boxdot B(x) \text{ dividido entre } G(x)^\ddagger$$

Es fácil verificar que $(F_n[x], \blacktriangle, \triangle)$ es un anillo, el cual se conoce como el *anillo de polinomios módulo G(x).*

---

f Supongamos que $a_i = 0$ para $i > n$ y $b_i = 0$ para $i > m$.

% La división de polinomios se lleva a cabo de la misma manera en que lo hicimos con el álgebra de la secundaria con, desde luego, la adición y la multiplicación como las operaciones de campo en $(F, +, \cdot)$.

$\triangle$	0	1	x	$1+x$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\triangle$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	1	$1+x$
$1+x$	0	$1+x$	$1+x$	0

Figura 11.21

Por ejemplo, sea $(F, +, \cdot)$ el campo de enteros módulo 2. Existen cuatro polinomios en $F_2[x]$, a saber, 0, 1, x, $1 + x$. Supongamos que $G(x)$ es $1 + x^2$. Las operaciones, $\triangle$ y $\triangle$ están definidas en la figura 11.21.

La construcción de anillos polinomiales conduce de inmediato a la construcción de campos de Galois. Un análisis detallado de este tópico se encuentra fuera del alcance de este libro.

Como ejemplo ilustrativo, consideremos una clase de códigos conocida como *códigos cíclicos.* Recordemos de nuestro análisis de la sección 11.7 que un código de bloque es una colección de palabras de la misma longitud. Un código de bloque es llamado un *código cíclico* si para toda palabra código $a_0a_1a_2 \cdots a_{n-2}a_{n-1}$, la secuencia $a_{n-1}a_0a_1a_2 \cdots a_{n-2}$ también es una palabra código. En otras palabras, para un código cíclico, un *corrimiento cíclico* de una palabra código también es una palabra código. Ahora mostraremos una manera muy compacta de describir códigos cíclicos.

Sea $(F, +, \cdot)$ el campo de los enteros módulo 2. Primero, señalemos que una sucesión binaria $a_0a_1a_2 \cdots a_{n-1}$ puede ser convenientemente representada por un polinomio

$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$

en el anillo polinomial $(F[x], \boxplus, \boxdot)$. Sea $(F_n[x], \triangle, \triangle)$ el anillo de polinomios módulo $1 + x^n$. Sea I un ideal de $(F_n[x], \triangle, \triangle)$. Queremos demostrar que los polinomios en / constituyen un código cíclico. Debido a que (/, A) es un grupo, los polinomios en / constituyen en efecto un código de grupo. Sea *A(x)* un polinomio en /. De acuerdo con la definición de un ideal, *x* A *A(x)* también es un polinomio en /. Sea

$$A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$

$$x \triangle A(x) = \text{residuo de } \frac{a_0x + a_1x^2 + \cdots + a_{n-1}x^n}{1 + x^n}$$

$$= \text{residuo de } \frac{a_0x + a_1x^2 + \cdots + a_{n-1} + a_{n-1}(1 + x^n)}{1 + x^n}$$

$$= a_{n-1} + a_0x + a_1x^2 + \cdots + a_{n-2}x^{n-1}$$

Por tanto, los polinomios en / constituyen un código cíclico.

Por otro lado, supongamos que / al conjunto de polinomios correspondientes a las palabras código en un código cíclico. Es evidente que (/, A) es un grupo abeliano. Si $A(x)$ está en I, $x \triangle A(x)$ está en I, $x^j \triangle A(x)$ está en I, y $b_jx^j \triangle A(x)$ está en I, debido a que los polinomios en / forman un código cíclico. De esto se tiene que $B(x) \triangle A(x)$ está en / para cualquier $B(x)$ en $F_n[x]$.

Se puede decir mucho más acerca de la descripción de códigos cíclicos como ideales en anillos polinomiales. Véase, por ejemplo, las referencias citadas en la sección 11.13.

11.13 NOTAS Y REFERENCIAS

Algunas referencias generales útiles sobre sistemas algebraicos son Birkhoff y MacLane [2], Cohn [3], Herstein [4], Paley y Weichel [8]. Véase el capítulo 4 de Knuth [5] para mayor información sobre cadenas de adición. Véase el capítulo 5 de Liu [7] sobre la teoría de conteo de Pólya, la cual generaliza el teorema de Burnside de la sección 11.6. Para un análisis más extenso sobre teoría de codificación algebraica, véase Berlekamp [1], Lin [6] y Peterson y Weldon [9].

1. Berlekamp, E. R.: *Algébrale Coding Theory,* McGraw-Hill Book Company, Nueva York, 1968.
2. Birkhoff, G. y S. MacLane: *A Survey ofModern Algebra,* 3° ed., Macmillan Company, Nueva York, 1965.
3. Cohn, P. M.: *Universal Algebra,* Harper and Row, Nueva York, 1965.
4. Herstein, I. N.: *Topics in Algebra,* Blaisdell Publishing Company, Waltham, Mass., 1964.
5. Knuth, D. E.: *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms,* Addison-Wesley Publishing Company, Reading, Mass., 1969.
6. Lin, S.: *An Introduction ofError-correcting Codes,* Prentice-Hall, Englewood Cliffs, N.J., 1970.
7. Liu, C. L.: *Introduction of Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
8. Paley, H. y P. M. Weichsel: *A First Course in Abstract Algebra,* Holt, Rinehart and Winston, Nueva York, 1966.
9. Peterson, W. W. y E. J. Weldon, Jr.: *Error-correcting Codes,* 2ª ed., MIT Press, Cambridge, Mass., 1972.

PROBLEMAS

11.1 Se $a * b = \text{máx}\,(a, b)$ es. Para cada uno de los siguientes casos determine si $a * b = \text{mín}\,(a, b + 2)$

a) $a * b = a + b + 3$)

d) $a * b = a + 2b$

11.2 Sea $\{A, \bullet)$ un sistema algebraico donde • es una operación binaria tal que, para cualesquiera *a* y b en A, $a \star b = a$.

a) Demuestre que • es una operación asociativa.

b) ¿Es posible que • sea una operación conmutativa?

11.3 Sea $(A, \star)$ un sistema algebraico tal que para todo a, b, c, d en A

$$a \star a = a$$
$$(a \star b) \star (c \star d) = (a \star c) \star (b \star d)$$

Demuestre que

$$a \star (b \star c) = (a \star b) \star (a \star c)$$

11.4 Sea Z_n el conjunto de enteros $\{0, 1, 2, \ldots, n-1\}$. Sea $\odot$ la operación binaria sobre Z_n tal que

$$a \odot b = \text{el residuo de } ab \text{ dividido entre } n$$

a) Construya la tabla de la operación $\odot$ para $n = 4$.
b) Demuestre que $(Z_n, \odot)$ es un semigrupo para todo n.

11.5 Sea $(A, *)$ un semigrupo. Sea a un elemento en A. Considere la operación binaria $\square$ sobre A tal que, para todo x y y en A,

$$x \square y = x * a * y$$

Demuestre que $\square$ es una operación asociativa.

11.6 Sea $(A, *)$ un semigrupo. Además, para todo a y b en A, si $a \neq b$, entonces $a * b \neq b * a$.

a) Demuestre que para todo a en A,

$$a * a = a$$

b) Demuestre que para todo a, b en A,

$$a * b * a = a$$

c) Demuestre que para todo a, b, c en A,

$$a * b * c = a * c$$

Sugerencia: observe que $a * b = b * a$ implica $a = b$.

11.7 Sea $(A, *)$ un semigrupo. Demuestre que para a, b, c en A, si $a * c = c * a$ y $b * c = c * b$, entonces $(a * b) * c = c * (a * b)$.

11.8 Sea $(\{a, b\}, *)$ un semigrupo donde $a * a = b$. Demuestre que:

a) $a * b = b * a$
b) $b * b = b$

11.9 Sea $(A, *)$ un semigrupo conmutativo. Demuestre que si $a * a = a$ y $b * b = b$, entonces $(a * b) * (a * b) = a * b$.

11.10 Sea $(A, *)$ un semigrupo. Demuestre que si A es un conjunto finito, existe un a en A tal que $a * a = a$.

11.11 Sea $(A, *)$ un semigrupo. Además, existe un elemento a en A tal que para todo x en A existen u y v en A que satisfacen la relación

$$a * u = v * a = x$$

Demuestre que existe un elemento identidad en A.

11.12 Sea $(A, *)$ un semigrupo y e una identidad izquierda. Además, para todo x en A existe un $\hat{x}$ en A tal que $\hat{x} * x = e$.

a) Demuestre que para todo a, b, c en A, si $a * b = a * c$, entonces $b = c$.
b) Demuestre que $(A, *)$ es un grupo, al demostrar que e es un elemento identidad.

Sugerencia: observe que $\hat{x} * x * \hat{x} * x = e$.

11.13 Sea $(A, \star)$ un sistema algebraico tal que para todo a, b en A

$$(a \star b) \star a = a$$

$$(a \star b) \star b = (b \star a) \star a$$

a) Demuestre que $a \star (a \star b) = a \star b$ para todo a y b.
b) Demuestre que $a \star a = (a \star b) \star (a \star b)$ para todo a y b.

c) Demuestre que $a \star a = b \star b$ para todo *ay b*.

d) Si *e* denota al elemento $a \star a$, demuestre que $e \star a = a$ y $a \star e = e$.

e) Demuestre que $a \star b = b \star a$ si y sólo si $a = b$.

/) Si *(A,* *) satisface adicionalmente la condición

$$a \star b = (a \star b) \star b$$

demuestre que • es idempotente y conmutativa.

11.14 Un *grupoide central* es un sistema algebraico I$(A, \star)$,donde • es una operación binaria tal que

$$(a \star b) \star (b \star c) = b$$

para todo *a,b,cenA.*

a) Demuestre que

$$a \star ((a \star b) \star c) = a \star b$$

$$(a \star (b \star c)) \star c = b \star c$$

en un grupoide central.

b) Sea $(A, \star)$ un sistema algebraico donde • es una operación binaria tal que

$$(a \star ((b \star c) \star d)) \star (c \star d) = c$$

para todo *a, b, c, d* en *A.* Demuestre que $(A, \star)$ es un grupoide central.

Sugerencia: utilice el resultado $(a \star ((d \star (b \star c)) \star d)) \star ((b \star c) \star d) = b \star c$ para demostrar que $(b \star c) \star (c \star d) = c$.

c) Construya un grafo dirigido con cuatro vértices en el cual exista exactamente un paseo de longitud 2 entre dos vértices cualesquiera.

d) Sea *G = (V, E)* un grafo dirigido en el cual existe exactamente un paseo de longitud 2 entre dos vértices cualesquiera. Para dos vértices cualesquiera *aybenV,* sean *(a, c)* y *(c, b)* las dos aristas en el paseo de *a* hacia *b.* Definimos un sistema algebraico $(V, \star)$ tal que $a \star b =$ *c.* Demuestre que $(V, \star)$ es un grupoide central.

e) Sea $(A, \star)$ un grupoide central. Construya un grafo dirigido $G = (A, E)$ tal que exista una arista *(a, c)* en £ si y sólo si existe un *b* en *A* tal que $a \star b = c$. Demuestre que existe exactamente un paseo de longitud 2 entre cualesquiera dos vértices en *A.*

/) Sean i$(A, \star)$ un grupoide central y *G = (A, E)* el correspondiente grafo dirigido, definido como en el inciso *e).* Para *a e A,* sean

$$R(a) = \{b|(a, b) \in E\}$$

$$L(a) = \{b|(b, a) \in E\}$$

Demuestre que para cualesquiera *a,benA, R(a)* y *L(b)* tienen el mismo número de elementos. Demuestre que para cualesquiera *a, b en A, R(á)* y *R(b)* tienen el mismo número de elementos. *g)* Sea *(A, -k)* un grupoide central. Demuestre que *A* tiene m^2 elementos para algún entero *m.*

Sugerencia: suponga que existen *m* elementos en *R{a).* ¿Qué puede decirse acerca de los conjuntos $R(b_1), R(b_2), \ldots, R(b_n)$ para $b_1, b_2, \ldots, b_m$ en $R(a)$?

11.15 Sea $(A, \star)$ un grupo. Demuestre que para cualesquiera i$a, b, c, d, a_1, b_1, c_1, d_1$ en A, si

$$a \star c = a_1 \star c_1$$

$$a \star d = a_1 \star d_1$$

$$b \star c = b_1 \star c_1$$

$$b \star d = b_1 \star d_1$$

entonces

$$b \star d = b_1 \star d_1$$

Sugerencia: observe que $b \star d = b \star (c \star c^{-1}) \star (a^{-1} \star a) \star d$.

11.16 *a)* Demuestre que cualquier grupo que contiene exactamente dos elementos es isomorfo a $(Z_2, \oplus)$.

b) Demuestre que cualquier grupo que contiene exactamente tres elementos es isomorfo a $(Z_3, \oplus)$.

c) ¿Cuántos grupos no isomorfos que contienen exactamente cuatro elementos existen?

11.17 Sean $(A, \star)$ y $(B, *)$ dos sistemas algebraicos. El producto cartesiano de $(A, \star)$ y $(B, *)$ es un sistema algebraico $(A \times B, \square)$, donde $\square$ es una operación binaria tal que p $b_1)$ y (a_2, b_2) en $A \times B$

$$(a_1, b_1) \square (a_2, b_2) = (a_1 \star a_2, b_1 * b_2)$$

Demuestre que el producto cartesiano de dos grupos es un grupo.

11.18 Sea $(A, \cdot)$ un grupo.

a) Demuestre que $(ab)^{-1} = b^{-1}a^{-1}$.

b) Demuestre que $(a_1a_2 \cdots a_{r-1}a_r)^{-1} = a_r^{-1}a_{r-1}^{-1} \cdots a_2^{-1}a_1^{-1}$.

c) Demuestre que $(b^ib^j)^{-1} = b^{-j}a^{-i}$; [$b^{-j}$ denota a $(b^{-1})^j$ y a^{-i} denota a $(a^{-1})^i$].

11.19 Sea i$(A, *)$ un monoide tal que para todo x en A, $x * x = e$, donde e es el elemento identidad. Demuestre que $(A, *)$ es un grupo abeliano.

11.20 Sea $(G, \star)$ un grupo y H un subconjunto no vacío de G. Demuestre que $(H, \star)$ es un subgrupo si para a y b cualesquiera en H, $a \star a^{-1}$ está también en H.

11.21 Demuestre que cualquier subgrupo de un grupo cíclico es también cíclico.

11.22 El *orden* de un elemento a en un grupo se define como el menor entero positivo m tal que $a^m = e$ (si ninguna potencia positiva de a es igual a e, el orden de a se define como infinito). Demuestre que, en un grupo finito, el orden de un elemento divide al orden del grupo.

11.23 Sean //, y H_2 subgrupos de un grupo G, tales que ninguno contiene al otro. Demuestre que existe un elemento de G que no pertenece a //, ni tampoco a H_2.

11.24 Sea $(H, \cdot)$ un subgrupo de $(G, \cdot)$. Sea $N = \{x | x \in G, xHx^{-1} = H\}$. Demuestre que $(N, \cdot)$ es un subgrupo de (G, •)•

11.25 Sea $(A, \star)$ un grupo donde A tiene un número par de elementos. Demuestre que existe un elemento a en A tal que $a \cdot a = e$, donde e es el elemento identidad y $a \neq e$.

11.26 Sealun $(A, \star)$ grupo. Sea B un subconjunto de A tal que Dem A, $a = b_1 \star b_2$ para algún b_1 y b_2 en B.

Sugerencia: considere el conjunto $C = \{a \star b^{-1} | b \in B\}$.

11.27 Sean $(H, \cdot)$ y $(K, \cdot)$ subgrupos del grupo (G, ■). Sea

$$HK = \{h \cdot k | h \in H, k \in K\}$$

Demuestre que (*HK*, ■) es un subgrupo de $(G, \cdot)$ si y sólo si $HK = KH$.

11.28 Sea $(A, \star)$ un grupo. Demuestre que $(A, *)$ es un grupo abeliano si y sólo si $a^2 \star b^2 = (a \star b)^2$ para todo *a* y *b* en *A*.

11.29 Sea $(A, \star)$ un grupo. Demuestre que $(A, \cdot)$ es un grupo abeliano si $a^3 \star b^3 = (a \star b)^3$, $a^4 \star b^4 = (a \star b)^4$, y $a^5 \star b^5 = (a \star b)^5$ para todo a y b en A.

11.30 Sea $(G, \star)$ un grupo de orden par. Sea (//,•) un subgrupo de $(G, \star)$ donde $|H| = |G|/2$. Demuestre que $(H, \star)$ es un subgrupo normal.

11.31 Sea $(H, \star)$ un subgrupo del grupo $(G, \star)$. Demuestre que $(H, \star)$ es un subgrupo normal si y sólo si $a \star H \star a^{-1} \subseteq H$ para todo $a \in G$.

11.32 Sea $(G, \star)$ un grupo. Sea $H = \{a | a \in G$ y $a \star b = b \star a$ para todo $b \in G\}$. Demuestre que H es un subgrupo normal.

11.33 *a)* Sean $(H, \star)$ y $(K, \star)$ subgrupos de un grupo $(G, \star)$. Demuestre que $(H \cap K, \star)$ también es un subgrupo.

b) Demuestre que, si $(H, \star)$ y $(K, \star)$ son subgrupos normales, entonces $(H \cap K, \star)$ también es un subgrupo normal.

11.34 Sean $(H, \star)$ y $(K, \star)$ subgrupos de un grupo $(G, \star)$. Demuestre que la función /desde H x K hacia G tal que $f[(h, k)] = h \star k$ para todo h en H y k en K es un isomorfismo desde el producto cartesiano de $(H, \star)$ y $(K, \star)$ hacia $(G, \star)$ si y sólo si:

1. H y K son subgrupos normales;
2. $\{h \star k | h \in H, k \in K\} = G$;
3. $H \cap K = \{e\}$, donde e es el elemento identidad.

11.35 Sea $(H, \star)$ un subgrupo normal del grupo $(G, \star)$. Demuestre que la imagen homomorfa de $(G, \star)$ inducida por el subgrupo $(H, \star)$ es un grupo abeliano si y sólo si $a \star b \star a^{-1} \star b^{-1}$ está en H para todo a y b en G.

11.36 Sean $(A, \star)$ y $(B, *)$ dos sistemas algebraicos y/un homomorfismo desde $(A, \star)$ hacia $(B, *)$.

a) Demuestre que, si $\star$ es una operación asociativa, también lo es *.

b) Demuestre que, si e es un elemento identidad en (A, *), entonces/(e) es un elemento identidad en $(B, *)$.

c) Demuestre que, si b es un inverso de a en $(A, \star)$, entonces $f(b)$ es un inverso de $f(a)$ en $(B, *)$.

d) Demuestre que una imagen homomorfa de un anillo es un anillo.

11.37 Sean/, y f_2 homomorfísmos desde un sistema algebraico $(A, \star)$ hacia otro sistema algebraico $(B, *)$. Sea g una función de A hacia B tal que

$$g(a) = f_1(a) * f_2(a)$$

para todo a en A. Demuestre que g es un homomorfismo de $(A, \star)$ hacia $(B, *)$ si $(B, *)$ es un semigrupo conmutativo.

11.38 Sean/y g homomorfísmos de un grupo $(G, \star)$ hacia un grupo $(H, \star)$. Demuestre que $(C, \star)$ es un subgrupo de $(G, \star)$, donde

$$C = \{x \in G | f(x) = g(x)\}$$

11.39 Una varilla dividida en seis segmentos se colorea con uno o más de cuatro colores diferentes. ¿De cuántas maneras se puede hacer esto?

11.40 ¿De cuántas maneras distintas pueden pintarse con tres colores los sectores del círculo de la figura 11P.1?

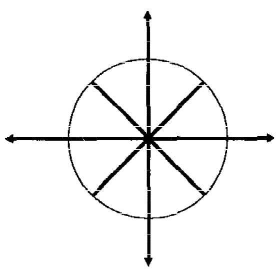


Figura 11P.1

11.41 a) Determine el número de tableros de ajedrez de 2 x 2 distintos cuyos cuadros están pintados en blanco y negro. Dos tableros de ajedrez se consideran distintos si uno no puede ser obtenido del otro mediante rotaciones.

b) Repita el inciso *a)* para tableros de ajedrez de 4 x 4.

11.42 *a)* Demuestre que para $p < \frac{1}{2}$, $(1-p)^{n-t_1} p^{t_1} > (1-p)^{n-t_2} p^{t_2}$ para $t_1 < t_2$.

b) Demuestre que para $p > \frac{1}{2}$, $(1-p)^{n-t_1} p^{t_1} < (1-p)^{n-t_2} p^{t_2}$ para $t_1 < t_2$.

11.43 *a)* Considere el código $G = \{00000, 11111\}$. Construya la tabla del conjunto cociente para mostrar que G de hecho puede corregir los errores de transmisión-simple y transmisión-doble.

b) Un código G contiene 16 palabras código: 0000000, 1111111, 1101000 y todos sus corrimientos cíclicos, 0010111 y todos sus corrimientos cíclicos. Demuestre que $(G, \oplus)$ es un código de grupo. Establezca la tabla del conjunto cociente que muestre que G puede corregir todos los errores de transmisión-simple.

11.44 Sea $(A, \star, *)$ un sistema algebraico con e_1 y e_2 los elementos identidad respecto de las operaciones $\star$ y $*$, respectivamente. Puesto que las operaciones $\star$ y $*$ son distributivas una sobre la otra, demuestre que $x \star x = x$ y $x * x = x$ para todo x en A.

Sugerencia: demuestre que $e_1 = e_1 * e_1$ y $e_2 = e_2 \star e_2$.

11.45 Sean $(A, \star, *)$ un sistema algebraico, donde

$$a \star b = a$$

para todo a, b en A, y * una operación binaria arbitraria. Demuestre que * es distributiva sobre ■*.

11.46 Sea $(A, +, \cdot)$ un anillo tal que $a \cdot a = a$ para todo a en A.

a) Demuestre que $a + a = 0$ para todo a, donde 0 es la identidad aditiva.

b) Demuestre que la operación • es conmutativa.

11.47 Demuestre que un dominio integral que tiene un número finito de elementos es un campo.

11.48 Un *cuadro latino de n* x *n* es un arreglo de tamaño *n* x *n* formado de *n* símbolos distintos de manera que ningún símbolo aparece mas de una vez en una fila o una columna. Dado un cuadro latino *A*, usaremos *a¡j* para denotar al símbolo en la ¡'-ésima fila y lay-ésima columna de *A*. Diremos que dos cuadros latinos *A* y *B* de *n* x *n* son *ortogonales* si no existen i, j, k, l tales que $a_{ij} = a_{kl}$ y $b_{ij} = b_{kl}$. Se dice que un conjunto de cuadros latinos es ortogonal si dos cualesquiera de ellos son ortogonales.

a) Construya un cuadro latino de 4 x 4.

b) Construya un par de cuadros latinos ortogonales de 3 x 3.

c) Sea $(\{b_0, b_1, \ldots, b_{n-1}\}, +, \cdot)$ un campo donde b_0 es la identidad aditiva. Construyamos los $(n-1)$ cuadrados de $n \times n$ $A_1, A_2, \ldots, A_{n-1}$ de tal manera que

$$a_{ij}^{(r)} = b_r \cdot b_i + b_j \qquad i, j = 0, 1, 2, \ldots, n-1$$
$$r = 1, 2, \ldots, n-1$$

donde $a_{ij}^{(r)}$ es la ij-ésima entrada del cuadro A_r. Demuestre que cada cuadro es un cuadro latino. Demuestre que el conjunto es un conjunto ortogonal.

11.49 Diremos que un anillo $(A, +, \cdot)$ es un *anillo conmutativo con la unidad* si $(A, \cdot)$ es un monoide conmutativo.

a) Diremos que un ideal //de un anillo es un *ideal primo* si, para dos elementos cualesquiera *a* y *b* que no están en *H*, *a* ■ *b* tampoco está en *H*. Demuestre que la imagen homomorfa de un anillo conmutativo con la unidad inducida por un ideal es un dominio integral si y sólo si el ideal es primo.

b) Diremos que un ideal *H* es un *ideal maximal* si los únicos ideales de un anillo que contiene a *H* son //y el anillo mismo. Demuestre que la imagen homomorfa de un anillo conmutativo con la unidad inducida por un ideal es un campo si y sólo si el ideal es maximal.

11.50 *a)* Sea $(F, +, \cdot)$ el campo de enteros módulo 2 y $(F[x], \boxplus, \square)$ el correspondiente anillo de polinomios. Construya el anillo de polinomios módulo $1 + x + x^2$.

ti) Sea $(F, +, \cdot)$ el campo de enteros módulo 3 y $(F[x], \boxplus, \square)$ el correspondiente anillo de polinomios. Construya el anillo de polinomios módulo $2 + 2x + x^2$.

11.51 Sea $(R, +, \cdot)$ el campo de los números reales, y sean $(R[x], \boxplus, \square)$ el correspondiente anillo de polinomios, y $(R_2[x], \blacktriangle, \Delta)$ el anillo de polinomios módulo $1 + x^2$.

a) Para $a + bx$ y $c + dx$ en $R_2[x]$, determine i$(a + bx) \blacktriangle (c + dx)$ y $(a + bx) \Delta (c + dx)$.

b) ¿Observa alguna similitud entre $(R_2, \blacktriangle, \Delta)$ y el campo de los números complejos?

CAPÍTULO

DOCE

Álgebras booleanas

12.1 LATTICES Y SISTEMAS ALGEBRAICOS

Recordemos que un lattice es un conjunto parcialmente ordenado en el cual dos elementos cualesquiera tienen una única cota superior mínima y una única cota inferior máxima. Por ejemplo, la figura 12.1 muestra un lattice. Existe una manera natural de definir un sistema algebraico con dos operaciones de manera que corresponda a un lattice dado. Sea $(A, \leq)$ un lattice. Definimos un sistema algebraico $(A, \vee, \wedge)$, donde $\vee$ y $\wedge$ son dos operaciones binarias sobre A tales que para a y b en A, $a \vee b$ es igual a la cota superior mínima de a y b, y $a \wedge b$ es igual a la cota inferior máxima de a y A. Nos referiremos a $(A, \vee, \wedge)$ como al sistema algebraico definido por el lattice $(A, \leq)$. Por ejemplo, en la figura 12.2 se muestra el sistema algebraico definido por el lattice de la figura 12.1. La operación binaria $\vee$ se conoce como la operación de *adición* y la operación binaria $\wedge$ se conoce como la operación de *multiplicación.* En consecuencia, la cota superior mínima de *a y b* también se reconocerá como la *adición* de *a* y *b,* y la cota inferior máxima de *a* y *b* se reconocerá como la *multiplicación* de *a* y *b.*

Por ejemplo, sea $\mathscr{P}(S)$ el conjunto potencia de un conjunto *S.* Recordemos que $(\mathscr{P}(S), \subseteq)$ es un lattice. Por tanto, esto define un sistema algebraico $(\mathscr{P}(S), \vee, \wedge)$. Para $S = \{a, b, c\}$, el lattice $(\mathscr{P}(S), \subseteq)$ se muestra en la figura 12.3*a* y las tablas para las operaciones $\vee$ y $\wedge$ se muestran en la figura 12.3*b*. Nos percatamos que la adición de dos subconjuntos de *S* es la unión de los dos subconjuntos y que la multiplicación de dos subconjuntos de *S* es la intersección de los dos subconjuntos. [Cuando introdujimos las nociones de unión e intersección de conjuntos en el capítulo 1, de manera intencional no abordamos el tópico de operaciones binarias. En efecto, todos los resultados de la sección 1.2 sobre la combinación de conjuntos para dar lugar a nuevos conjuntos pueden ser convenientemente parafraseados en términos de operaciones binarias sobre $\mathscr{P}(S)$.]

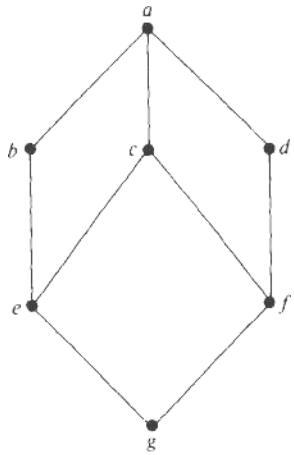


Figura 12.1

Otro ejemplo, consideremos el conjunto parcialmente ordenado $(N, \leqq)$, donde N es el conjunto de los números naturales y $\leqq$ es la relación "menor o igual que" entre números naturales. Uno puede verificar rápidamente que $(N, \leqq)$ es un lattice, y que éste define un sistema algebraico $(N, \vee, \wedge)$ tal que $a \vee b = \text{máx}\ (a, b)$ y $a \wedge b = \text{mín}\ (a, b)$.

Del mismo modo, como otro ejemplo, sean N^+ el conjunto de todos los enteros positivos, y | una relación binaria sobre N^+ tal que $a|b$ si y sólo si a divide a b. Dejamos al lector verificar que $(N^+, |)$ es un lattice. Más aún, la adición de dos elementos a y b es el mínimo común múltiplo de a y b, y la multiplicación de dos elementos a y b es el máximo común divisor de a y b.

Tenemos los siguientes resultados:

Teorema 12.1 Para todo a y b en un lattice $(A, \leq)$

$$a \leq a \vee b \tag{12.1}$$
$$a \wedge b \leq a \tag{12.2}$$

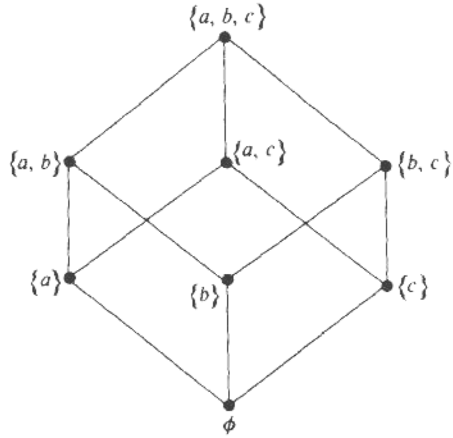
DEMOSTRACIÓN Debido a que la adición de a y b es una cota superior de a, $a \leq a \vee b$. Debido a que la multiplicación de a y b es una cota inferior de a, $a \wedge b \leq a$.

❐

$\vee$	a	b	c	d	e	f	g
a	a	a	a	a	a	a	a
b	a	b	a	a	b	a	b
c	a	a	c	a	c	c	c
d	a	a	a	d	a	d	d
e	a	b	c	a	e	c	e
f	a	a	c	d	c	f	f
g	a	b	c	d	e	f	g

$\wedge$	a	b	c	d	e	f	g
a	a	b	c	d	e	f	g
b	b	b	e	g	e	g	g
c	c	e	c	f	e	f	g
d	d	g	f	d	g	f	g
e	e	e	e	g	e	g	g
f	f	g	f	f	g	f	g
g	g	g	g	g	g	g	g

Figura 12.2



a)

$\vee$	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$	ϕ
$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$
$\{a, b\}$	$\{a, b, c\}$	$\{a, b\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b\}$	$\{a, b\}$	$\{a, b, c\}$	$\{a, b\}$
$\{a, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, c\}$	$\{a, b, c\}$	$\{a, c\}$	$\{a, b, c\}$	$\{a, c\}$	$\{a, c\}$
$\{b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{b, c\}$	$\{a, b, c\}$	$\{b, c\}$	$\{b, c\}$	$\{b, c\}$
$\{a\}$	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{a, b, c\}$	$\{a\}$	$\{a, b\}$	$\{a, c\}$	$\{a\}$
$\{b\}$	$\{a, b, c\}$	$\{a, b\}$	$\{a, b, c\}$	$\{b, c\}$	$\{a, b\}$	$\{b\}$	$\{b, c\}$	$\{b\}$
$\{c\}$	$\{a, b, c\}$	$\{a, b, c\}$	$\{a, c\}$	$\{b, c\}$	$\{a, c\}$	$\{b, c\}$	$\{c\}$	$\{c\}$
ϕ	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$	ϕ

$\wedge$	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$	ϕ
$\{a, b, c\}$	$\{a, b, c\}$	$\{a, b\}$	$\{a, c\}$	$\{b, c\}$	$\{a\}$	$\{b\}$	$\{c\}$	ϕ
$\{a, b\}$	$\{a, b\}$	$\{a, b\}$	$\{a\}$	$\{b\}$	$\{a\}$	$\{b\}$	ϕ	ϕ
$\{a, c\}$	$\{a, c\}$	$\{a\}$	$\{a, c\}$	$\{c\}$	$\{a\}$	ϕ	$\{c\}$	ϕ
$\{b, c\}$	$\{b, c\}$	$\{b\}$	$\{c\}$	$\{b, c\}$	ϕ	$\{b\}$	$\{c\}$	ϕ
$\{a\}$	$\{a\}$	$\{a\}$	$\{a\}$	ϕ	$\{a\}$	ϕ	ϕ	ϕ
$\{b\}$	$\{b\}$	$\{b\}$	ϕ	$\{b\}$	ϕ	$\{b\}$	ϕ	ϕ
$\{c\}$	$\{c\}$	ϕ	$\{c\}$	$\{c\}$	ϕ	ϕ	$\{c\}$	ϕ
ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ

b)

Figura 12.3

Teorema 12.2 Para todo *a*, *b*, *c*, *d* en un lattice $(A, \leq)$, si

$$a \leq b \qquad \text{y} \qquad c \leq d$$

entonces

$$a \vee c \leq b \vee d \tag{12.3}$$
$$a \wedge c \leq b \wedge d \tag{12.4}$$

DEMOSTRACIÓN Puesto que

$$b \leq b \vee d \quad d \leq b \vee d$$

por transitividad

$$a \leq b \vee d \quad c \leq b \vee d$$

En otras palabras, $b \vee d$ es una cota superior de a y c. Y ya que $a \vee c$ es la cota superior mínima de a y c, tenemos que†

$$a \vee c \leq b \vee d$$

Puesto que

$$a \wedge c \leq a \qquad a \wedge c \leq c$$

por transitividad

$$a \wedge c \leq b \qquad a \wedge c \leq d$$

En otras palabras, $a \wedge c$ es una cota inferior de b y d. Y ya que $b \wedge d$ es la cota inferior máxima de b y d, tenemos que

$$a \wedge c \leq b \wedge d$$

❐

12.2 PRINCIPIO DE DUALIDAD

Este principio de dualidad es un concepto importante y aparece en muchos lugares y contextos diferentes. Comencemos con algunos sencillos ejemplos ilustrativos. En Estados Unidos, todos los automóviles de pasajeros tienen el asiento del conductor en la parte frontal izquierda. En consecuencia, se siguen reglas de tránsito como las siguientes:

Los automóviles siempre circulan por el lado derecho de una calle.
Sobre una autopista de varios carriles, los carriles de la izquierda son carriles para rebasar.
Los automóviles pueden dar vuelta a la derecha en un semáforo en rojo.

Sin embargo, en Inglaterra todos los automóviles de pasajeros tienen el asiento del conductor en la parte frontal derecha, así, se siguen reglas de tránsito como las siguientes:

Los automóviles siempre circulan por el lado izquierdo de una calle.
Sobre una autopista de varios carriles, los carriles de la derecha son carriles para rebasar.
Los automóviles pueden dar vuelta a la izquierda en un semáforo en rojo.

Observamos cómo los conceptos de izquierda y derecha pueden ser intercambiados en estas dos situaciones de manera que reglas de tránsito significativas en Estados Unidos se conviertan en reglas de tránsito significativas en Inglaterra, y viceversa. Decimos que los conceptos de izquierda y derecha son conceptos duales.

---

† Recordamos al lector que, por definición, en un conjunto parcialmente ordenado, una cota superior de dos elementos es una cota superior mínima o mayor que una cota superior mínima.

Otro ejemplo, tenemos una historia que adaptamos libremente de la obra de ficción china *Flores en el espejo!* Había un país llamado La Tierra de los Caballeros. La siguiente conversación es un ejemplo común del comportamiento de sus pobladores:

Cliente:	Quiero comprar este vaso. Es el más hermoso.
Vendedor:	El vaso en realidad no parece tan bueno. Por favor note que tiene algunos defectos en el trabajo de pintura.
Cliente:	Usted no llamaría a esto defectos, ¿o lo haría? El vaso es una pieza perfecta de artesanía. ¿Cuánto quiere usted por él?
Vendedor:	Por lo común vendería un vaso como éste en \$ 10. Puesto que usted ha sido uno de nuestros mejores clientes por años, lo venderé en \$15.
Cliente:	\$ 15 es muy poco. No puedo permitir aprovecharme de usted, ¿qué tal \$20?
Vendedor:	Yo obtendría una buena ganancia a \$15. Pero si usted insiste, hagamos un trato en \$16.
Cliente:	Mi esposa se quejará de que no pago lo suficiente por las cosas que compro. No obstante, estoy dispuesto a comprar el vaso en \$18.

Es interesante el hecho de que muchos de los conceptos del negocio de regateo en La Tierra de los Caballeros son opuestos a los que estamos más acostumbrados. En efecto, es muy obvio cómo podemos refrasear la conversación entre el cliente y el vendedor en La Tierra de los Caballeros para hacer que nos suene más familiar, y, recíprocamente, cómo podemos refrasear una conversación entre un cliente y un vendedor que con frecuencia escuchamos para hacer que ésta suene más familiar a los personajes de La Tierra de los Caballeros, mediante la adecuada sustitución de los conceptos de precios altos y bajos, transacciones de negocios redituables y no redituables, buena y mala calidad del trabajo, etcétera. De nuevo, estos pares de conceptos son pares de conceptos duales.

Esperamos que el lector se encuentre listo para un análisis más serio. Sea $(A, \leq)$ un conjunto parcialmente ordenado. Sea $\leq_R$ una relación binaria sobre *A* tal que para *a y b* en $(A, \leq_R)$,*b* si y sólo si $b \leq a$. No es difícil notar que $(A, \leq_R)$ también es un conjunto parcialmente ordenado. Más aún, si $(A, \leq)$ es un lattice, también lo es $(A, \leq_R)$.‡ Observamos que los lattices $(A, \leq)$ y $(A, \leq_R)$ están muy relacionados, al igual que lo están los sistemas algebraicos definidos por ellos. Para ser específicos, la operación de adición del sistema algebraico definido por $(A, \leq)$ es la operación de multiplicación del sistema algebraico definido por $(A, \leq_R)$, y la operación de multiplicación del sistema algebraico definido por $(A, \leq)$ es la operación de adición del sistema algebraico definido por $(A, \leq_R)$. En consecuencia, dada una proposición válida cualquiera referente a las propiedades generales de los lattices podemos obtener otra proposición valida reemplazando la relación $\leq$ por $\geq$,§ el término *operación de*

† El título original es *Ching Hua Yuan.* Escrito por Li Ju-Chen (1763-1830). Una traducción al inglés por Lin Tai-Yi fue publicada por University of California Press, en 1965.

‡ Véase el problema 4.27.

§ Es claro, $a \leq_R b$ si y sólo si $a \geq b$.

adición por el término *operación de multiplicación,* y el término *operación de multiplicación* por el termino *operación de adición.* Esto es conocido como el *principio de dualidad para lattices.*

Como ejemplo, veamos que la relación en (12.1) puede establecerse como, "la adición de dos elementos cualesquiera en un lattice es mayor o igual que cada uno de los dos elementos". En correspondencia, tenemos que "la multiplicación de dos elementos cualesquiera en un lattice es menor o igual que cada uno de los dos elementos ", lo cual es la relación en (12.2). Asimismo, de acuerdo con el principio de dualidad, la relación en (12.4) puede obtenerse a partir de la relación en (12.3), y viceversa. Estudiaremos más ejemplos sobre la aplicación del principio de dualidad conforme procedamos a mostrar algunas de las propiedades generales de los sistemas algebraicos definidos como lattices.†

12.3 PROPIEDADES BÁSICAS DE SISTEMAS ALGEBRAICOS DEFINIDAS POR LATTICES

Ahora mostraremos algunas de las propiedades básicas que poseen los sistemas algebraicos definidos como lattices. Sea $(A, \vee, \wedge)$ el sistema algebraico definido por un lattice $(A, \leq)$. Tenemos:

Teorema 12.3 Las operaciones de adición y multiplicación son conmutativas.

DEMOSTRACIÓN Ésta se obtiene directamente de la definición para la cota superior mínima y la cota inferior máxima para dos elementos en un lattice.

❐

Teorema 12.4 Las operaciones de adición y multiplicación son asociativas.

DEMOSTRACIÓN Primero mostraremos que la operación de adición v es asociativa. Esto es, para *a, b, c* en *A,*

$$a \vee (b \vee c) = (a \vee b) \vee c$$

Sea

$$a \vee (b \vee c) = g$$

y

$$(a \vee b) \vee c = h$$

Puesto que g es la adición de a y $(b \vee c)$,

$$a \leq g \qquad\qquad b \vee c \leq g$$

† Existen otros ejemplos importantes que ilustran el principio de dualidad. En ingeniería eléctrica los conceptos de voltaje y corriente, resistencia y conductividad, inductancia y capacitancia son conceptos duales (para mayores detalles véase, por ejemplo, [6]). En teoría de grafos, los conceptos de circuitos y conjuntos de corte, árbol y árbol-cociente son conceptos duales (para más detalles véase, por ejemplo, [8]).

Además, $b \vee c \leq g$ significa que

$$b \leq g \qquad c \leq g$$

Puesto que la adición de *a* y *b* es la cota superior mínima de *a y b*, a partir de $a \leq g$ y $b \leq g$ obtenemos

$$a \vee b \leq g$$

lo cual aunado a que $c \leq g$, nos conduce a que

$$(a \vee b) \vee c \leq g$$

Así hemos demostrado que

$$h \leq g$$

De manera similar, podemos demostrar que

$$g \leq h$$

En consecuencia, debido a la propiedad de antisimetría de una relación de orden parcial, concluimos que

$$g = h$$

De acuerdo con el principio de dualidad, la operación de multiplicación $\wedge$ también es asociativa.

□

Teorema 12.5 Para todo a en A, $a \vee a = a$ y $a \wedge a = a$.

DEMOSTRACIÓN De acuerdo con (12.1),

$$a \leq a \vee a$$

Y ya que $a \leq a$, obtenemos que†

$$a \vee a \leq a$$

De esto se tiene que

$$a \vee a = a$$

De acuerdo con el principio de dualidad, también tenemos que

$$a \wedge a = a$$

□

Los resultados del teorema 12.5 son conocidos como la *propiedad de idempotencia* de las operaciones de adición y multiplicación.

Teorema 12.6 Para todo *a* y *b* en *A*,

$$a \vee (a \wedge b) = a$$
$$a \wedge (a \vee b) = a$$

† Si $g \leq f$ y $h \leq f$, entonces $g \vee h \leq f$ debido a que $g \vee h$ es la cota superior mínima de g y h.

DEMOSTRACIÓN Puesto que $a \vee (a \wedge b$ es la adición de a y $a \wedge b$, tenemos que

$$a \leq a \vee (a \wedge b) \tag{12.5}$$

Y ya que

$$a \leq a \qquad a \wedge b \leq a$$

de acuerdo con (12.3), tenemos que

$$a \vee (a \wedge b) \leq a \vee a \tag{12.6}$$

Puesto que

$$a \vee a = a$$

(12.6) se convierte en

$$a \vee (a \wedge b) \leq a \tag{12.7}$$

Combinando (12.5) y (12.7), obtenemos

$$a \vee (a \wedge b) = a$$

Que $a \wedge (a \vee b) = a$ se obtiene del principio de dualidad. ❐

Los resultados del teorema 12.6 son conocidos como la *propiedad de absorción* de las operaciones de adición y multiplicación.

Por ejemplo, consideremos el lattice de los números naturales ordenados por la relación "menor o igual que", $(N, \leqq)$. Debido a que el máximo de dos números *a* y *b* es el mismo que el máximo de los dos números *b* y *a,* la operación de adición en $(N, \vee, \wedge)$ es conmutativa. De manera análoga, la operación de multiplicación también es conmutativa. Debido a que tanto máx (máx (a, *b), c)* como máx (a, máx (6, c)) son iguales al máximo de los tres números *a, b y c,* la operación de adición en $(N, \vee, \wedge)$ es asociativa. Asimismo, la operación de multiplicación es asociativa. La propiedad de idempotencia para las operaciones de adición y multiplicación se sigue del hecho de que el máximo de a y a es a, y de que el mínimo de a y a es a. Las ecuaciones

máx [*a*, mín *(a, b)*] = *a*
mín [a, máx *(a, b)*] = *a*

dan lugar a la propiedad de absorción.

Como otro ejemplo, consideremos el lartice de los enteros positivos ordenados por la relación "es divisible por", $(N^+, |)$. Debido a que el mínimo común múltiplo de *a* y *b* es el mismo que el mínimo común múltiplo de *b* y a, la operación de adición en $(N^+, \vee, \wedge)$ es conmutativa. También la operación de multiplicación es conmutativa. Debido a que tanto $(a \vee b) \vee c$ como $a \vee (b \vee c)$ son el mínimo común múltiplo de los tres números *a, b* y c, la operación de adición es asociativa. De la misma manera, la operación de multiplicación es asociativa. El mínimo común divisor de a y a es a, como lo es el máximo común divisor de

a y a. Por tanto, se cumple la propiedad de idempotencia de las operaciones de adición y multiplicación. Las ecuaciones

$$\text{mcm}\,[a, \text{mcd}\,\{a, b\}] = a$$
$$\text{mcd}\,[a, \text{mcm}\,\{a, b\}] = a$$

dan lugar a la propiedad de absorción.

12.4 LATTICES DISTRIBUTIVOS Y COMPLEMENTADOS

Ahora estudiaremos clases de lattices que poseen propiedades adicionales. Naturalmente, podemos anticipar que estos lattices definirán sistemas algebraicos que están aún más "estructurados". Se dice que un lattice es un lattice *distributivo* si la operación de multiplicación se distribuye sobre la operación de adición y la operación de adición se distribuye sobre la operación de multiplicación. Esto es, para todo a,b y c,

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$
$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)^{\dagger}$$

El lattice de la figura 12.3*a* es un lattice distributivo. No obstante, el lattice de la figura 12.4 no es un lattice distributivo. Observemos que en el lattice de la figura 12.4,

$$b \wedge (c \vee d) = b \wedge a = b$$
$$(b \wedge c) \vee (b \wedge d) = e \vee e = e$$

Nuestra definición de lattice distributivo es ligeramente redundante como se muestra en el teorema 12.7.

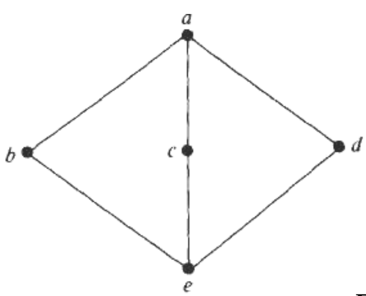


Figura 12.4

† Debido a la conmutatividad, están implicadas las ecuaciones

$$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$$
$$(a \wedge b) \vee c = (a \vee c) \wedge (b \vee c)$$

Teorema 12.7

Si la operación de multiplicación es distributiva sobre la operación de adición en un lattice, entonces la operación de adición también es distributiva sobre la operación de multiplicación. Si la operación de adición es distributiva sobre la operación de multiplicación, entonces la operación de multiplicación también es distributiva sobre la operación de adición.

DEMOSTRACIÓN Puesto que

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

obtenemos que

$$\begin{aligned}(a \vee b) \wedge (a \vee c) &= [(a \vee b) \wedge a] \vee [(a \vee b) \wedge c] \\ &= a \vee [(a \vee b) \wedge c] \\ &= a \vee [(a \wedge c) \vee (b \wedge c)] \\ &= [a \vee (a \wedge c)] \vee (b \wedge c) \\ &= a \vee (b \wedge c)\end{aligned}$$

Por dualidad, obtenemos el resultado de que si la operación de adición $\vee$ es distributiva sobre la operación de multiplicación $\wedge$, entonces la operación de multiplicación $\wedge$ también es distributiva sobre la operación de adición $\vee$.

Un elemento *a* en un lattice $(A, \leq)$ se conoce como una *cota inferior universal* si para todo elemento $b \in A$, $a \leq b$. Por ejemplo, para el lattice mostrado en la figura 12.5, *h* es una cota inferior universal. Se sigue a partir de la definición que si un lattice posee una cota inferior universal, ésta es única. Puesto que si suponemos que existen dos cotas inferiores universales *a* y *b*, entonces

$$a \leq b \qquad \text{y} \qquad b \leq a$$

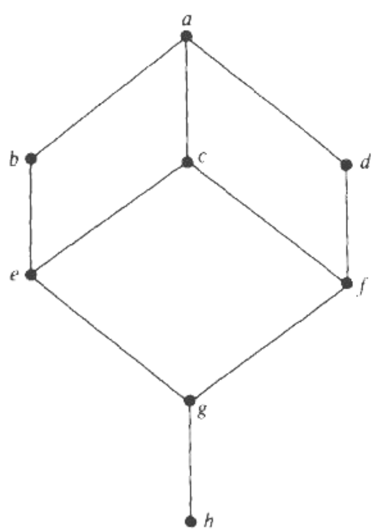


Figura 12.5

implican

$$a = b$$

Un elemento a en un lattice $(A, \leq)$ se conoce como una *cota superior universal* si para todo elemento $b \in A$, $b \leq a$. De nuevo, si un lattice posee una cota superior universal, ésta es única.

Usaremos 0 para denotar a la cota inferior universal y 1 para denotar a la cota superior universal de un lattice (si tales cotas existen). Señalamos que 0 es en efecto la identidad de la operación de adición $\vee$, y que 1 es la identidad de la operación de multiplicación $\wedge$. Por ejemplo, en el lattice $(\mathscr{P}(S), \subseteq)$ donde $\mathscr{P}(S)$ es el conjunto potencia del conjunto S y $\subseteq$ es la relación de inclusión, el conjunto vacío $\varnothing$ es la cota inferior universal y el conjunto S es la cota superior universal. Así, tenemos:

Teorema 12.8 Sea $(A, \leq)$ un lattice con cotas inferior y superior universales 0 y 1. Para todo elemento a en A,

$$a \vee 1 = 1 \qquad a \wedge 1 = a$$
$$a \vee 0 = a \qquad a \wedge 0 = 0$$

DEMOSTRACIÓN De acuerdo con (12.1), $1 \leq a \vee 1$. Puesto que 1 es la cota superior universal, $a \vee 1 \leq 1$. Así, $a \vee 1 = 1$.

De acuerdo con (12.2), $a \wedge 1 \leq a$. Puesto que $a \leq a$ y $a \leq 1$, de acuerdo con (12.4), $a \wedge a \leq a \wedge 1$, o $a \leq a \wedge 1$. Así, $a \wedge 1 = a$.

Las otras dos relaciones pueden demostrarse en forma similar. ☐

Sea $(A, \leq)$ un lattice con cotas inferior y superior universales 0 y 1. Para un elemento a en A, se dice que un elemento b es un *complemento* de a si

$$a \vee b = 1 \quad \text{y} \quad a \wedge b = 0$$

Observemos que debido a la conmutatividad, si a es un complemento de b, entonces b es también un complemento de a. Por ejemplo, para el lattice de la figura 12.1, d es un complemento de e y e es un complemento de d. Observemos que un elemento podría tener más de un complemento. Por ejemplo, en el lattice de la figura 12.1, tanto b como e son complementos de d. No obstante, por otro lado, en el lattice de la figura 12.1, c no tiene complemento. Sin embargo, 0 es el único complemento de 1 y 1 es el único complemento de 0.†

Diremos que un lattice es un *lattice complementado* si todo elemento en el lattice posee un complemento (es evidente, un lattice complementado debe poseer cota inferior y superior universales). Por ejemplo, el lattice de la figura 12.6 es un lattice complementado, donde el complemento de a es c, el complemento de b también es c, y de esto se sigue que los complementos de c son a y b.

El teorema 12.9 muestra que la distributividad es una propiedad muy "útil".

† Es claro que 0 es un complemento de 1, y 1 es un complemento de 0. Dejamos al lector demostrar su unicidad (véase el problema 12.14).

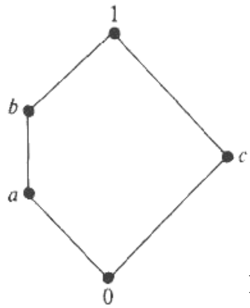


Figura 12.6

Teorema 12.9 En un lattice distributivo, si un elemento posee un complemento entonces este complemento es único.

DEMOSTRACIÓN Supongamos que un elemento *a* posee dos complementos *b* y *c*. Esto es,

$$a \vee b = 1 \qquad a \wedge b = 0$$
$$a \vee c = 1 \qquad a \wedge c = 0$$

Luego tenemos que

$$\begin{aligned} b &= b \wedge 1 \\ &= b \wedge (a \vee c) \\ &= (b \wedge a) \vee (b \wedge c) \\ &= 0 \vee (b \wedge c) \\ &= (a \wedge c) \vee (b \wedge c) \\ &= (a \vee b) \wedge c \\ &= 1 \wedge c \\ &= c \end{aligned}$$

❐

12.5 LATTICES BOOLEANOS Y ÁLGEBRAS BOOLEANAS

Un lattice complementado y distributivo también se conoce como un *lattice booleano.* Sea $(A, \leq)$ un lattice booleano. Debido a que todo elemento en un lattice booleano tiene un complemento único, podemos definir una operación unaria sobre *A*, denotada por $^{-}$, de manera que para todo a en A $\overline{a}$ es el complemento de *a.* † La operación unaria $^{-}$ se conoce como la operación de *complementación.* En consecuencia, diremos que un lattice booleano $(A, \leq)$ define un sistema algebraico $(A, \vee, \wedge, ^{-})$, donde $\vee$, $\wedge$, y $^{-}$ son las operaciones de

† En lugar de escribir –*a* la notación $\overline{a}$ es más conveniente. Así, por ejemplo, el complemento de la adición de *a* y *b* se escribirá como $\overline{a \vee b}$.

adición, multiplicación y complementación, respectivamente. Un sistema algebraico definido por un lattice booleano se conoce como *álgebra booleana*.

Podemos construir ejemplos de álgebras booleanas con facilidad. Sea S un conjunto finito. Recordemos que $(\mathscr{P}(S), \subseteq)$ es un lattice. Además, $(\mathscr{P}(S), \subseteq)$ es un lattice booleano en el cual la cota superior universal es S, la cota inferior universal es $\varnothing$, y el complemento de cualquier conjunto T en $\mathscr{P}(S)$ es el conjunto $S - T$. Ahora demostraremos el:

Teorema 12.10 Para todo a y b en un álgebra booleana

$$\overline{a \vee b} = \bar{a} \wedge \bar{b}$$
$$\overline{a \wedge b} = \bar{a} \vee \bar{b}$$

DEMOSTRACIÓN Tenemos que

$$\begin{aligned}(a \vee b) \vee (\bar{a} \wedge \bar{b}) &= [(a \vee b) \vee \bar{a}] \wedge [(a \vee b) \vee \bar{b}] \\ &= [(a \vee \bar{a}) \vee b] \wedge [a \vee (b \vee \bar{b})] \\ &= 1 \wedge 1 \\ &= 1\end{aligned}$$

y

$$\begin{aligned}(a \vee b) \wedge (\bar{a} \wedge \bar{b}) &= [a \wedge (\bar{a} \wedge \bar{b})] \vee [b \wedge (\bar{a} \wedge \bar{b})] \\ &= [(a \wedge \bar{a}) \wedge \bar{b}] \vee [(b \wedge \bar{b}) \wedge \bar{a}] \\ &= 0 \vee 0 \\ &= 0\end{aligned}$$

Así, $\bar{a} \wedge \bar{b}$ es el complemento de $a \vee b$. Esto es,

$$\overline{a \vee b} = \bar{a} \wedge \bar{b}$$

Que $\overline{a \wedge b} = \bar{a} \vee \bar{b}$ se sigue del principio de dualidad. ❐

Los resultados del teorema 12.10 son conocidos como leyes de DeMorgan.

12.6 UNICIDAD DE LAS ALGEBRAS BOOLEANAS FINITAS

Uno podría imaginarse que existe bastante libertad para construir multitudes de álgebras booleanas diferentes. Pero resulta que en realidad éste no es el caso. Demostraremos que un álgebra booleana finita† tiene exactamente 2^n elementos para algún $n > 0$. Además, existe un álgebra booleana única de 2^n elementos para todo $n > 0$.

Sean a y b dos elementos en un lattice. Recordemos que se dice que a es una *cubierta*

† El álgebra booleana finita tiene un número finito de elementos.

de b si $b < a$[†] y no existe un elemento c tal que $b < c$ y $c < a$. Sea $(A, \leq)$ un lattice con una cota inferior universal 0. Diremos que un elemento es un *átomo* si éste es una cubierta de 0. Por ejemplo, para el lattice de la figura 12.1, e y f son átomos.

Sea $(A, \leq)$ un lattice finito con una cota inferior universal. Afirmamos que para todo elemento b diferente de cero,[‡] existe al menos un átomo a tal que $a \leq b$. Es evidente que si b es un átomo, esto es inmediato. Si b no es un átomo, debido a que $(A, \leq)$ es un lattice finito, debe existir una cadena en $(A, \leq)$ tal que $0 < b_i < \ldots < b_2 < b_1 < b$, donde b_i es un átomo. Y de esto se sigue que $b_i \leq b$.

Ahora demostraremos una secuencia de lemas:

Lema 12.1 En un lattice distributivo, si $b \wedge \overline{c} = 0$, entonces $b \leq c$.

DEMOSTRACIÓN Puesto que $b \wedge \overline{c} = 0$, tenemos que

$$(b \wedge \overline{c}) \vee c = c$$

De acuerdo con la ley distributiva, tenemos

$$(b \vee c) \wedge (\overline{c} \vee c) = c$$

o

$$b \vee c = c \tag{12.8}$$

Puesto que (12.8) significa que c es la cota superior mínima de b y c, $b \leq c$. ❐

Lema 12.2 Sea $(A, \vee, \wedge, \overline{\ })$ un álgebra boolena finita. Si b es cualquier elemento diferente de cero en A, y $a_1, a_2, \ldots, a_k$ son todos los átomos de A tales que $a_i \leq b$, entonces $b = a_1 \vee a_2 \vee \cdots \vee a_k$.

DEMOSTRACIÓN Puesto que

$$a_1 \leq b \qquad a_2 \leq b \qquad \cdots \qquad a_k \leq b$$

de esto se sigue de inmediato que

$$a_1 \vee a_2 \vee \cdots \vee a_k \leq b$$

Por conveniencia notacional, denotemos por c a $a_1 \vee a_2 \vee \cdots \vee a_k$.

Supongamos que $b \wedge \overline{c} \neq 0$. En tal caso, existe un átomo a tal que $a \leq (b \wedge \overline{c})$. Puesto que

$$b \wedge \overline{c} \leq b$$

y

$$b \wedge \overline{c} \leq \overline{c}$$

de acuerdo con la ley transitiva, tenemos que

$$a \leq b \tag{12.9}$$

† Usamos la notación $b < a$ para indicar que $b \leq a$ y $b \neq a$.
‡ Por un elemento diferente de cero queremos decir un elemento que no es igual a la cota inferior universal 0.

y

$$a \leq \overline{c}$$

De acuerdo con (12.9), *a* es igual a alguno de los átomos $a_1, a_2, \ldots, a_k$. De lo que se sigue que

$$a \leq c$$

Al combinar $a \leq \overline{c}$ y $a \leq c$, obtenemos

$$a \leq c \wedge \overline{c}$$

o

$$a \leq 0$$

lo cual es imposible.

De esto se sigue que debemos tener $b \wedge \overline{c} = 0$. En ese caso, de acuerdo con el lema 12.1, $b \leq c$. Por la ley de antisimetría, tenemos que

$$b = a_1 \vee a_2 \vee \cdots \vee a_k$$

❐

Lema 12.3 Sea $(A, \vee, \wedge, \overline{\ })$ un álgebra booleana finita. Si *b* es un elemento cualquiera distinto de cero en A, y $a_1, a_2, \ldots, a_k$ son todos los átomos de *A* tales que $a_i \leq b$. Entonces $b = a_1 \vee a_2 \vee \cdots \vee a_k$ es la *única* manera de representar a *b* como una adición de átomos.

DEMOSTRACIÓN Supongamos que tenemos una representación alternativa

$$b = a_{j_1} \vee a_{j_2} \vee \cdots \vee a_{j_t}$$

Es claro, ya que *b* es la cota superior mínima de $a_{j_1}, a_{j_2}, \ldots, a_{j_t}$

$$a_{j_1} \leq b \qquad a_{j_2} \leq b \cdots a_{j_t} \leq b$$

En otras palabras, si *b* es expresado como una adición de átomos, éstos deberán ser átomos menores o iguales a *b*.

Ahora consideremos un átomo a_{j_u}, $1 \leqq u \leqq t$. Puesto que $a_{j_u} \leq b$, tenemos

$$a_{j_u} \wedge b = a_{j_u}$$

Esto es,

$$a_{j_u} \wedge (a_1 \vee a_2 \vee \cdots \vee a_k) = a_{j_u}$$

o

$$(a_{j_u} \wedge a_1) \vee (a_{j_u} \wedge a_2) \vee \cdots \vee (a_{j_u} \wedge a_k) = a_{j_u}$$

Entonces, para algún a_i, $1 \leqq i \leqq k$,

$$a_{j_u} \wedge a_i \neq 0$$

Debido a que tanto a_{j_u} como a_i son átomos, debemos tener que $a_{j_u} = a_i$.

Así, cada átomo en la representación alternativa es un átomo en la representación original, de lo cual se cumple el lema.

❐

Se sigue de los lemas 12.1,12.2 y 12.3 que existe una correspondencia uno a uno entre los elementos de un lattice booleano y los subconjuntos de los átomos. De hecho, esta correspondencia uno a uno es un isomorfísmo de $(A, \leq)$ hacia $(\mathscr{P}(S), \subseteq)$, donde S es el conjunto de átomos. Así tenemos:

Teorema 12.11 Sea $(A, \vee, \wedge, ^-)$ un álgebra booleana finita. Sea S el conjunto de átomos. Entonces $(A, \vee, \wedge, ^-)$ es isomorfo al sistema algebraico definido por el lattice $(\mathscr{P}(S), \subseteq)$.

Se sigue del teorema 12.11 que existe un álgebra booleana única y finita de 2" elementos para cualquier $n > 0$. Además, no existen otras álgebras booleanas finitas.

12.7 FUNCIONES BOOLEANAS Y EXPRESIONES BOOLEANAS

Sea $(A, \vee, \wedge, ^-)$ un álgebra booleana. Consideremos las funciones de A^n hacia A. Por ejemplo, la tabla de la figura 12.7 muestra una función/de $\{0, 1\}^3$ hacia $\{0, 1\}$, y la tabla de la figura 12.8 muestra una función/de $\{0, 1, 2, 3\}^2$ hacia $\{0, 1, 2, 3\}$. Aunque una función siempre puede ser descrita exhaustivamente en forma tabular, estamos interesados en formas alternativas para describir funciones. Recordemos que existe la posibilidad de especificar una función mediante una "expresión de forma cerrada". Tratemos de alcanzar efectivamente tal posibilidad.

Sea $(A, \vee, \wedge, ^-)$ un álgebra booleana. Una *expresión booleana* sobre $(A, \vee, \wedge, ^-)$ se define de la siguiente manera:†

1. Cualquier elemento de A es una expresión booleana.
2. Cualquier nombre de variable es una expresión booleana.
3. Si e_1 y e_2 son expresiones booleanas, entonces $\bar{e}_1$, $e_1 \vee e_2$, $e_1 \wedge e_2$ son expresiones booleanas.

Por ejemplo,

$$0 \vee x$$
$$((\overline{2 \wedge 3}) \vee (x_1 \vee \bar{x}_2)) \wedge (\overline{x_1 \wedge x_3})$$

son expresiones booleanas sobre el álgebra booleana $(\{0, 1, 2, 3\}, \vee, \wedge, ^-)$. Una expresión booleana que contiene n variables *distintas* se conoce como una expresión booleana de n variables.

Sea $E(x_1, x_2, \ldots, x_n)$ una expresión booleana de n variables sobre un álgebra booleana $(A, \vee, \wedge, ^-)$. Por una asignación de valores a las variables $x_1, x_2, \ldots, x_n$, queremos indicar la asignación de elementos de A como los valores de las variables. Para una asignación de

† Véase la sección 11.12 para la definición de una expresión algebraica sobre un sistema algebraico.

	f
(0, 0, 0)	0
(0, 0, 1)	0
(0, 1, 0)	1
(0, 1, 1)	0
(1, 0, 0)	1
(1, 0, 1)	1
(1, 1, 0)	0
(1, 1, 1)	1

Figura 12.7

	f
(0, 0)	1
(0, 1)	0
(0, 2)	0
(0, 3)	3
(1, 0)	1
(1, 1)	1
(1, 2)	0
(1, 3)	3
(2, 0)	2
(2, 1)	0
(2, 2)	1
(2, 3)	1
(3, 0)	3
(3, 1)	0
(3, 2)	0
(3, 3)	2

Figura 12.8

valores a las variables podemos evaluar la expresión $E(x_1, x_2, \ldots, x_n)$ mediante la sustitución de las variables en la expresión por sus valores. Por ejemplo, para la expresión booleana

$$E(x_1, x_2, x_3) = (x_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_2) \wedge (\overline{x_2 \vee x_3})$$

sobre el álgebra booleana $(\{0, 1\}, \vee, \wedge, {}^{-})$ la asignación de valores $x_1 = 0, x_2 = 1, x_3 = 0$ da lugar a

$$\begin{aligned} E(0, 1, 0) &= (0 \vee 1) \wedge (\bar{0} \vee \bar{1}) \wedge (\overline{1 \vee 0}) \\ &= 1 \wedge 1 \wedge 0 \\ &= 0 \end{aligned}$$

Diremos que dos expresiones booleanas de *n* variables son *equivalentes* si tienen el mismo valor para cualquier asignación de valores para las *n* variables. Por ejemplo, el lector puede verificar rápidamente que las dos expresiones booleanas

$$(x_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_3)$$
$$x_1 \wedge (x_2 \vee \bar{x}_3)$$

son equivalentes. Luego escribimos

$$E_1(x_1, x_2, \ldots, x_n) = E_2(x_1, x_2, \ldots, x_n)$$

para indicar que las dos expresiones $E_1(x_1, x_2, \ldots, x_n)$ y $E_2(x_1, x_2, \ldots, x_n)$ son equivalentes. Así, cuando decimos que manipulamos o simplificamos una expresión booleana siempre queremos decir en realidad que la manipulamos o simplificamos hacia una forma equivalente. Debido a que los elementos de *A* serán asignados como valores de las variables en una expresión booleana, todas las igualdades que derivemos en las secciones previas que involucran elementos de un álgebra booleana, pueden ser aplicadas para manipular y simplificar expresiones booleanas. Por ejemplo, observemos que

$$
\begin{aligned}
(x_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_3) &= x_1 \wedge (x_2 \vee \bar{x}_3) \\
(x_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_2) &= x_1 \\
x_1 \wedge x_2 &= (x_1 \wedge x_2) \wedge 1 \\
&= (x_1 \wedge x_2) \wedge (x_3 \vee \bar{x}_3) \\
&= (x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge \bar{x}_3)
\end{aligned}
$$

No es difícil imaginar cómo debemos especificar una función de *A″* hacia *A* mediante una expresión booleana $E(x_1, x_2, \ldots, x_n)$. A saber, hacemos que cada asignación de valores a las variables $x_1, x_2, \ldots, x_n$, sea una *n*-ada ordenada en el dominio *A″*, y hacemos que el correspondiente valor de $E(x_1, x_2, \ldots, x_n)$ sea la imagen correspondiente en el rango *A.* Por ejemplo, el lector puede verificar inmediatamente que la expresión booleana

$$(\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3) \vee (x_1 \wedge \bar{x}_2) \vee (x_1 \wedge x_3)$$

sobre el álgebra booleana $(\{0, 1\}, \vee, \wedge, {}^{-})$ define la función/en la figura 12.7.

Por otro lado, podemos preguntamos si es posible que cualquier función de *A″* hacia *A* puede ser especificada mediante una expresión booleana sobre $(A, \vee, \wedge, {}^{-})$. La respuesta es negativa. Por ejemplo, no existe expresión booleana sobre el álgebra de cuatro elementos que define la función de la figura 12.8 (véase problema 12.24). Una función de *A″* hacia *A* se denomina *función booleana* si puede ser especificada mediante una expresión booleana (de *n* variables).

Señalemos, sin embargo, que para el caso del álgebra booleana bivaluada cualquier función de {0,1}" hacia {0,1} es una función booleana. En efecto, mostraremos dos maneras para obtener una expresión booleana que especifique una función dada de {0, 1}" hacia {0, 1}. Diremos que una expresión booleana de *n* variables *x-¡*, x_2,. . ., x_n es un *minterm* si ésta es de la forma

$$\tilde{x}_1 \wedge \tilde{x}_2 \wedge \cdots \wedge \tilde{x}_n \tag{12.10}$$

donde usamos $\tilde{x}_i$ para denotar ya sea x_i o bien $\bar{x}_i$. Diremos que una expresión booleana sobre $(\{0, 1\}, \vee, \wedge, {}^{-})$ está en su *forma normal disyuntiva* si es una adición de minterms. Por ejemplo,

$$(\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3) \vee (\bar{x}_1 \wedge x_2 \wedge \bar{x}_3) \vee (x_1 \wedge x_2 \wedge x_3) \tag{12.11}$$

es una expresión booleana en forma normal disyuntiva. Más aún, existen tres minterms en la expresión, a saber, $\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3$, $\bar{x}_1 \wedge x_2 \wedge \bar{x}_3$ y $x_1 \wedge x_2 \wedge x_3$.

Diremos que una expresión booleana de *n* variables $x_1, x_2, \ldots, x_n$ es un *maxterm* si es de la forma

$$\tilde{x}_1 \vee \tilde{x}_2 \vee \cdots \vee \tilde{x}_n \tag{12.12}$$

donde, nuevamente usamos $\tilde{x}_i$ para denotar ya sea x_i o bien $\bar{x}_i$. Se dice que una expresión booleana sobre $(\{0, 1\}, \vee, \wedge, {}^{-})$ está en *su forma normal conjuntiva* si es una multiplicación de maxterms. Por ejemplo,

$$(x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_1 \vee x_2 \vee \bar{x}_3) \wedge (\bar{x}_1 \wedge \bar{x}_2 \vee x_3) \tag{12.13}$$

es una expresión booleana en forma normal conjuntiva consistente de cinco maxterms.

Dada una función de {0,1)" a {0,1}, podemos obtener una expresión booleana en forma normal disyuntiva correspondiente a esta función, si hacemos corresponder un minterm a cada w-ada ordenada de números 0 y números 1 para las cuales el valor de la función es 1. Específicamente, para cada una de tales *n*-adas, tenemos un minterm

$$\tilde{x}_1 \wedge \tilde{x}_2 \wedge \cdots \wedge \tilde{x}_n$$

en el cual $\tilde{x}_i$ es x_i si la *z*-ésima componente de la *n*-ada es 1, y essi $\bar{x}_i$ la *i*-ésima componente de la *n*-ada es 0.† Por ejemplo, la expresión booleana de (12.11) corresponde a la función *f* en la figura 12.9.

De modo similar, dada una función de {0, 1}" hacia {0, 1}, podemos obtener una expresión booleana en forma normal conjuntiva correspondiente a esta función, si hacemos corresponder un maxterm a cada *n*-ada ordenada de números 0 y números 1 para las cuales el valor de la función es 0. En específico, para cada una de tales *n*-adas, tenemos un maxterm

$$\tilde{x}_1 \vee \tilde{x}_2 \vee \cdots \vee \tilde{x}_n$$

en el cual $\tilde{x}_i$ es x_i si la *i*-ésima componente de la *n*-adas es 0, y es $\bar{x}_i$ la *i*-ésima componente de la *n*-adas es 1. Por ejemplo, la expresión booleana de (12.13) corresponde a la función *f* en la figura 12.9.

			f
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Figura 12.9

† Dejamos al lector convencerse de que la expresión booleana obtenida efectivamente especifica a la función dada.

$\vee$	F	T
F	F	T
T	T	T

$\wedge$	F	T
F	F	F
T	F	T

	$^{-}$
F	T
T	F

Figura 12.10

12.8 CALCULO PROPOSICIONA1

Recordemos nuestro análisis de la sección 1.8 sobre proposiciones: una proposición es un enunciado que bien puede ser verdadero (*T*) o ser falso *(F).* Más aún, las proposiciones se pueden combinar para dar lugar a nuevas proposiciones. En particular, introdujimos las nociones de disyunción, conjunción y negación de proposiciones. Ahora mostraremos que nuestro análisis puede formularse en el marco de referencia de un sistema algebraico.

Consideremos un sistema algebraico $(\{F, T\}, \vee, \wedge, ^{-})$ donde las definiciones de las Operaciones $\vee, \wedge$ y $^{-}$ se muestran en la figura 12.10. El lector puede verificar inmediatamente que $(\{F, T\}, \vee, \wedge, ^{-})$ es un álgebra booleana de dos elementos. Dentro de tal marco de referencia algebraico, una proposición atómica es una variable que puede tomar bien el valor *F* (falso) o el valor *T* (verdadero). Una tautología corresponde a la constante *T* y una contradicción corresponde a la constante *F*. La disyunción de dos proposiciones *p* y *q* puede ser representada mediante la expresión algebraica $p \vee q$. Observemos que la definición de la operación binaria v en la figura 12.10 es consistente con la definición de la disyunción de dos proposiciones que se mostró en la tabla de verdad de la figura 1.9. De modo similar, la conjunción de dos proposiciones *p* y *q* puede ser representada por la expresión algebraica $p \wedge q$, y la negación de la proposición *p* puede ser representada por la expresión algebraica $\bar{p}$. De nuevo, las definiciones de las operaciones $\wedge$ y $^{-}$ son consistentes con aquellas de la conjunción y la negación de proposiciones que se mostró en la figura 1.9. De esto se sigue que una proposición compuesta puede representarse por una expresión booleana. Además, la tabla de verdad de una proposición compuesta es exactamente la descripción tabular del valor de la correspondiente expresión booleana para todas las posibles combinaciones de los valores de las proposiciones atómicas. Concluimos así, que todos nuestros resultados sobre la manipulación y simplificación de expresiones booleanas pueden ser aplicados para manipular y simplificar proposiciones compuestas como se ilustra en los siguientes ejemplos.

Ejemplo 12.1

Consideremos el enunciado, "iré al juego de pelota si no hay examen mañana o bien si hay examen mañana y el juego es un encuentro de campeonato". Sean *p* la proposición "hay examen mañana", y *q* la proposición "el juego es un encuentro de campeonato". Es obvio que iré al juego de pelota si la proposición

$$\bar{p} \vee (p \wedge q)$$

es verdadera. No obstante, la igualdad

$$\bar{p} \vee (p \wedge q) = \bar{p} \vee q$$

nos permite simplificar nuestro enunciado a, "iré al juego de pelota si no hay examen mañana o si el juego es un encuentro de campeonato".

□

Ejemplo 12.2

Consideremos las siguientes instrucciones dadas a un técnico de mantenimiento:

1. La energía eléctrica deberá encenderse si no es el caso de que no haya alguien en la oficina y el sistema de monitoreo automático no esté en operación.
2. El sistema de monitoreo automático estará en operación si y sólo si no hay alguien en la oficina o una nómina grande se deja en la oficina.

Si *p* denota la proposición, "no hay alguien en la oficina", *q* denota la proposición "el sistema de monitoreo automático está en operación", y *r* denota la proposición "una nómina grande se deja en la oficina"; evidentemente la energía eléctrica se encenderá si la proposición

$$\overline{p \wedge \bar{q}}$$

es verdadera. No obstante, ya que

$$q = p \vee r$$

tenemos que

$$\begin{aligned} \overline{(p \wedge \bar{q})} &= \overline{p \wedge (\overline{p \vee r})} \\ &= \overline{p \wedge (\bar{p} \wedge \bar{r})} \\ &= T \end{aligned}$$

En consecuencia, podemos concluir que las instrucciones son completamente superfluas y que pueden ser remplazadas por la simple instrucción de que siempre debe dejarse la energía eléctrica encendida.

□

Ejemplo 12.3

Consideremos el grafo dirigido de la figura 12.11. Queremos determinar un subconjunto mínimo de aristas las cuales, si son eliminadas, destruirían todos los circuitos dirigidos en el grafo (por un conjunto *mínimo* de aristas, entendemos un conjunto tal que la eliminación de cualquier subconjunto propio de éste no destruirá todos los circuitos dirigidos). Observamos que en el grafo de la figura 12.11 existen cinco circuitos dirigidos, a saber, *(a, d, g,f, b)*, *(a, e,f, b)*, (c, *d, g,f)*, (c, *e,f)* y *(g, i, h)*. A fin de destruir el circuito dirigido (*a, d, g,f, b)*, debemos eliminar al menos una de las aristas *a, d, g, f, b*. Para destruir el circuito dirigido *(a, e,f, b)*, debemos eliminar al menos una de las

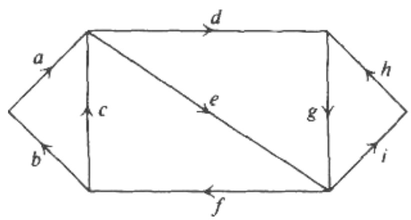


Figura 12.11

$\rightarrow$	F	T
F	T	T
T	F	T

$\leftrightarrow$	F	T
F	T	F
T	F	T

Figura 12.12

aristas *a, e,f, b*. En consecuencia, para destruir los cinco circuitos dirigidos, debemos eliminar al menos una arista de cada uno de ellos. A partir de la expresión algebraica,

$$(a \vee d \vee g \vee f \vee b) \wedge (a \vee e \vee f \vee b) \wedge (c \vee d \vee g \vee f) \wedge (c \vee e \vee f) \wedge (g \vee i \vee h)$$
$$= (a \wedge d \wedge e \wedge h) \vee (e \wedge g) \vee \cdots$$

la eliminación del subconjunto $\{a, d, e, h\}$, o $\{e, g\}, \ldots$ destruirá todos los circuitos dirigidos en el grafo.

❐

Por último, recordemos que definimos en la sección 1.8 dos maneras adicionales para combinar proposiciones, a saber, $p \rightarrow q$ y $p \leftrightarrow q$ para las proposiciones p y q dadas. Es claro que podemos incrementar nuestro sistema algebraico al definir las dos operaciones $\rightarrow$ y $\leftrightarrow$ como se muestra en la figura 12.12. No obstante, ese incremento es innecesario ya que siempre podremos remplazar expresiones algebraicas con las operaciones $\rightarrow$ y $\leftrightarrow$ por expresiones algebraicas equivalentes que utilicen únicamente las operaciones $\vee$, $\wedge$ y $^-$. En específico, a partir de las tablas de verdad para $p \rightarrow q$ y $p \leftrightarrow q$, obtenemos que

$$p \rightarrow q = \overline{p} \vee q$$
$$p \leftrightarrow q = (p \wedge q) \vee (\overline{p} \wedge \overline{q})$$

para las expresiones booleanas correspondientes a las proposiciones compuestas $p \rightarrow q$ y $p \leftrightarrow q$.†

Como un ejemplo, señalemos que

$$\begin{aligned} p \rightarrow q &= \overline{p} \vee q \\ &= \overline{\overline{q}} \vee \overline{p} \\ &= \overline{q} \rightarrow \overline{p} \end{aligned}$$

Por tanto, las proposiciones, "si la temperatura está por arriba de 30°C, iremos a la playa", "bien la temperatura no está por arriba de 30°C, o bien iremos a la playa", y "si no vamos a la playa, la temperatura no está por arriba de 30°C" son equivalentes. Dé modo similar, las proposiciones, "si un grupo es cíclico, es abeliano", "bien un grupo no es cíclico, o bien es abeliano", y "si un grupo no es abeliano, no es cíclico" son equivalentes. Como otro ejemplo, señalemos que

$$\begin{aligned} p \leftrightarrow q &= (\overline{p} \vee q) \wedge (p \vee \overline{q})^{\ddagger} \\ &= (p \rightarrow q) \wedge (q \rightarrow p) \end{aligned}$$

† La situación es similar conceptualmente, aquella de no definir la sustracción como otra operación en un anillo, puesto que la sustracción de un elemento es equivalente a la adición del inverso aditivo del elemento.

‡ Ésta es la forma normal conjuntiva para la expresión booleana correspondiente al enunciado compuesto $p \leftrightarrow q$.

Por tanto, el enunciado "habrá un cargo de $2 por servicio si y sólo si el balance mensual de la cuenta de cheques es menor que $100" es equivalente al enunciado, "si hay un cargo de $2 por servicio entonces debe ser el caso de que el balance mensual de la cuenta de cheques es menor que $100, y si el balance mensual de la cuenta de cheques es menor que $100 entonces habrá un cargo de $2 por servicio".

En este punto recordemos una observación hecha en la sección 1.8, referente a que existe una gran similitud entre la composición de proposiciones para formar nuevas proposiciones y la composición de conjuntos para formar nuevos conjuntos. En efecto, ya que varias de las propiedades y resultados sobre conjuntos y operaciones de conjuntos y sobre proposiciones y operaciones de proposiciones son aquellas de las álgebras booleanas, dicha similitud no es una coincidencia.

12.9 DISEÑO E IMPLANTACIÓN DE CIRCUITOS DIGITALES

Supongamos que queremos diseñar un circuito electrónico tal que accionará un sonido de alerta en un automóvil si la temperatura de la máquina excede los 200°F o si el automóvil se encuentra en la posición de avance y el conductor no ha abrochado su cinturón de seguridad. Es obvio que tenemos la relación

$$b = p \vee (q \wedge \overline{r})$$

donde *b* es la proposición "accionar el sonido de alerta",/? es la proposición "la temperatura de la máquina excede los 200°C", *q* es la proposición "el automóvil está en la posición de avance", y *r* es la proposición "el cinturón de seguridad del conductor está abrochado". Para construir un circuito electrónico que se comporte como hemos descrito, primero debemos decidir sobre una convención para representar las proposiciones mediante señales eléctricas. Por ejemplo, podríamos representar una proposición por un voltaje eléctrico. Si la proposición es verdadera, ésta será representada por un voltaje alto (digamos 6 V), y si la proposición es falsa, será representada por un bajo voltaje (digamos 0 V). Así, en el presente ejemplo, correspondiendo a la proposición *p*, aparecerá una señal de alto voltaje si la temperatura de la máquina excede los 200°F, y aparecerá una señal de bajo voltaje si la temperatura de la máquina no excede los 200°F. Del mismo modo, correspondiendo a la proposición *b*, una señal de alto voltaje, la cual puede ser usada para accionar el sonido de alerta, aparecerá si se cumplen las condiciones para accionar el sonido de alerta, y aparecerá una señal de bajo voltaje, la cual no será suficiente para accionar el sonido de alerta, si no se cumplen las condiciones para accionar el sonido de alerta. Una vez que hemos decidido sobre una convención para representar las proposiciones mediante señales eléctricas, podemos diseñar circuitos eléctricos que correspondan a las operaciones $\vee$, $\wedge$, y $\overline{\ }$. En el resto de esta sección tomaremos la convención de representar las proposiciones mediante voltajes, como se describió con anterioridad.

Una *compuerta-OR* es un circuito que tiene dos entradas y una salida como se muestra esquemáticamente en la figura 12.13*a*. El voltaje de salida de una compuerta-OR es alto si bien uno o ambos voltajes de entrada son altos, y el voltaje de salida es bajo si ambos voltajes de entrada son bajos. Es claro que la señal de salida de una compuerta-OR corresponde a

Compuerta OR

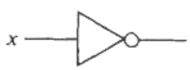
x	y	Salida
L	L	L
L	H	H
H	L	H
H	H	H

a)

Compuerta AND

x	y	Salida
L	L	L
L	H	L
H	L	L
H	H	H

b)



NOT gate

Compuerta NOT

x	Salida
L	L
H	L

c)

Figura 12.13

una proposición la cual es la disyunción de las proposiciones correspondientes a las señales de entrada.

Una *compuerta-AND* es un circuito que tiene dos entradas y una salida como se muestra esquemáticamente en la figura 12.13*b*. El voltaje de salida de una compuerta-AND es alto si ambos voltajes de entrada son altos, y el voltaje de salida es bajo si uno o ambos voltajes de entrada son bajos. Claro que la señal de salida de una compuerta-AND corresponde a una proposición la cual es la conjunción de las proposiciones correspondientes a las señales de entrada.

Una *compuerta-NOT*, o un *inversor*, es un circuito que tiene una entrada y una salida, como se muestra esquemáticamente en la figura 12.13c. Su voltaje de salida es alto si el voltaje de entrada es bajo, y su voltaje de salida es bajo si el voltaje de entrada es alto. La

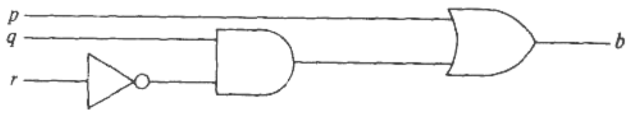


Figura 12.14

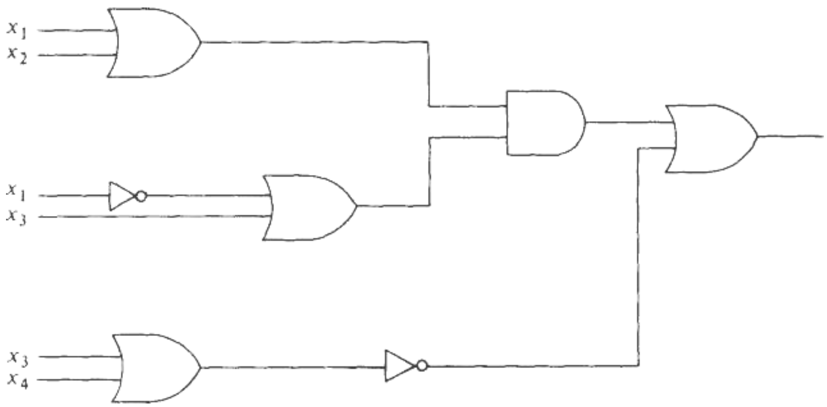


Figura 12.15

señal de salida de una compuerta-NOT corresponde a una proposición que es la negación de la proposición correspondiente a la señal de entrada.

Es obvio que podemos interconectar estos dispositivos para formar un circuito electrónico que realice cualquier expresión booleana dada. Por ejemplo, el circuito de la figura 12.14 realiza la expresión booleana

$$p \vee (q \wedge \overline{r})$$

correspondiente al problema del sonido de alerta del automóvil presentado inicialmente. Como otro ejemplo, el circuito de la figura 12.15 realiza la expresión booleana

$$[(x_1 \vee x_2) \wedge (\overline{x}_1 \vee x_3)] \vee \overline{(x_3 \vee x_4)}$$

Esperamos que nuestro breve análisis haya ilustrado las ideas esenciales en lo referente al diseño e implantación de redes digitales. No es difícil visualizar que el comportamiento de dispositivos computacionales y de control complejos pueden ser especificados como proposiciones compuestas y entonces realizados con componentes electrónicas. Exhortamos al lector para que busque un tratamiento más profundo y detallado sobre el diseño e implantación de redes digitales en textos sobre teoría de interruptores y diseño lógico.

12.10 CIRCUITOS DE INTERRUPTORES

Ahora mostraremos otro ejemplo de aplicación de álgebras booleanas. Consideremos un interruptor eléctrico simple de encendido-apagado que puede colocarse en la posición de

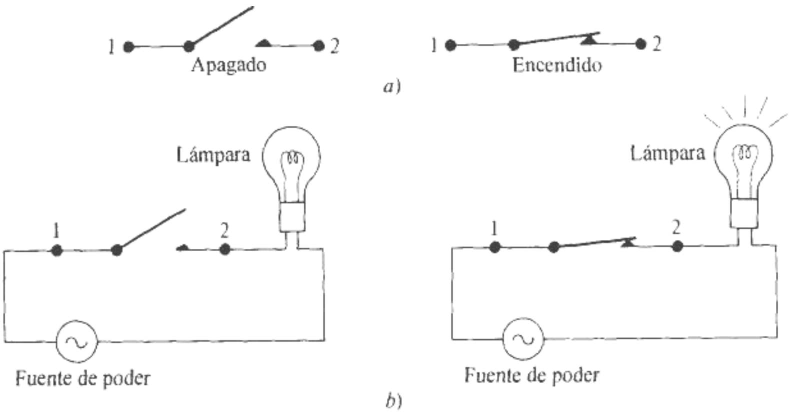


Figura 12.16

apagado o en la posición de *encendido*, como se muestra en la figura 12.16a. Cuando el interruptor está en la posición de apagado decimos que hay un *circuito abierto* entre las dos terminales 1 y 2. Cuando el interruptor está en la posición de encendido decimos que hay un *circuito cerrado* entre las dos terminales 1 y 2. Como es conocido por el lector, podemos utilizar un interruptor de encendido-apagado para encender o apagar cualquier dispositivo eléctrico, como se ilustra en la figura 12.16*b*.

Los interruptores pueden ser interconectados de varias maneras para realizar funciones de control complejas. Por ejemplo, un sistema de aire acondicionado podría ser controlado por los termostatos en dos habitaciones. Si la temperatura en cualquier habitación excede los 25C, el termostato en esa habitación cerrará un interruptor para encender el sistema de aire acondicionado. Como otro ejemplo, la puerta de una caja fuerte de un banco podría ser controlada electrónicamente de manera que no pueda ser abierta a menos que dos llaves separadas sean utilizadas para cerrar dos interruptores eléctricos de modo simultáneo.

Las dos formas más fundamentales de interconectar interruptores son la conexión en paralelo y la conexión en serie. Dos interruptores pueden ser interconectados en *paralelo*, como se muestra en la figura 12.17*a*, y en *serie*, como se muestra en la figura 12.17*b*. Es claro que en una conexión en paralelo de dos interruptores, existe un circuito abierto entre las terminales 1 y 2 si ambos interruptores están en la posición de apagado, y hay un circuito cerrado entre las terminales 1 y 2 si uno o ambos interruptores están en la posición de

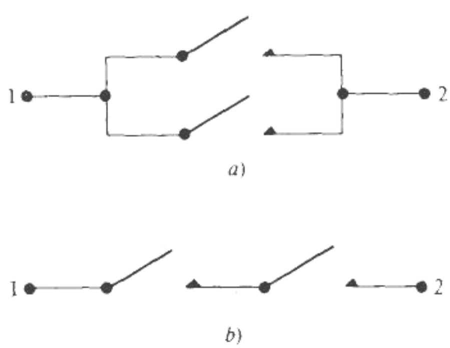


Figura 12.17

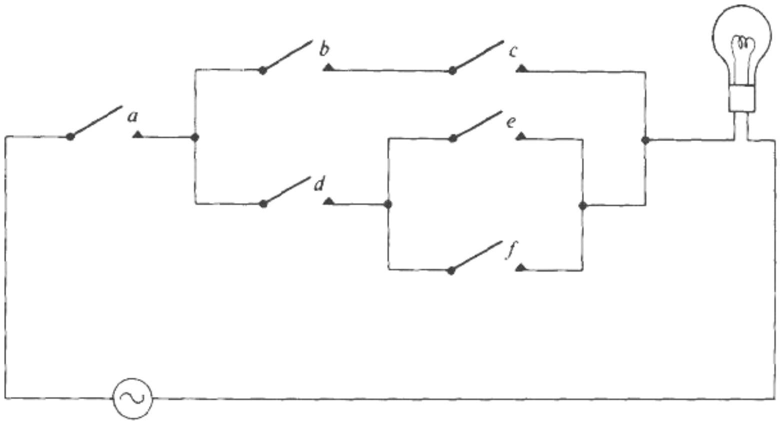


Figura 12.18

encendido. De modo similar, en una conexión en serie de dos interruptores, existe un circuito abierto entre las terminales 1 y 2 si uno o ambos interruptores están en la posición de apagado, y hay un circuito cerrado entre las terminales 1 y 2 si ambos interruptores están en la posición de encendido. Así, en el ejemplo del sistema de aire acondicionado, conectaremos los dos interruptores controlados por termostatos en paralelo. Por otro lado, en el ejemplo de la caja fuerte de un banco, conectaremos los dos interruptores controlados por las dos llaves en serie.

Debido a que podemos ver una conexión en paralelo o una conexión en serie de dos interruptores como un "super" interruptor que está ya sea abierto o cerrado, podemos interconectarlos con otros interruptores para formar redes de interruptores complejas, tales como la mostrada en la figura 12.18, donde se utilizan nombres simbólicos para denotar diferentes interruptores.

Uno se da cuenta de inmediato que la lámpara dé la red de la figura 12.18 puede ser encendida mediante diferentes asignaciones de posiciones para los interruptores de la red. Por ejemplo, la lámpara se encenderá si los interruptores *a, b, c* están cerrados, o si los interruptores *a, d, f* están cerrados, etcétera. Por otro lado, la lámpara se apagará si el interruptor *a* está abierto o si los interruptores *b, e, f,* están abiertos, etcétera. Para estudiar el comportamiento de redes de interruptores complejas, definiremos ahora un sistema algebraico. Sea ({abierto, cerrado}, **PAR, SER**) un sistema algebraico donde **PAR** y **SER** son dos operaciones binarias cuyas definiciones se muestran en la figura 12.19. Es claro que en este sistema algebraico, un interruptor es una variable que puede tomar ya sea el valor *abierto* o bien el valor *cerrado.* La operación binaria **PAR** corresponde a la conexión en paralelo de dos interruptores, y la operación binaria **SER** corresponde a la conexión en serie de dos interruptores. En consecuencia, una red de interruptoreo consistente de conexiones en

PAR	abierto	cerrado
abierto	abierto	cerrado
cerrado	cerrado	cerrado

SER	abierto	cerrado
abierto	abierto	abierto
cerrado	abierto	cerrado

Figura 12.19

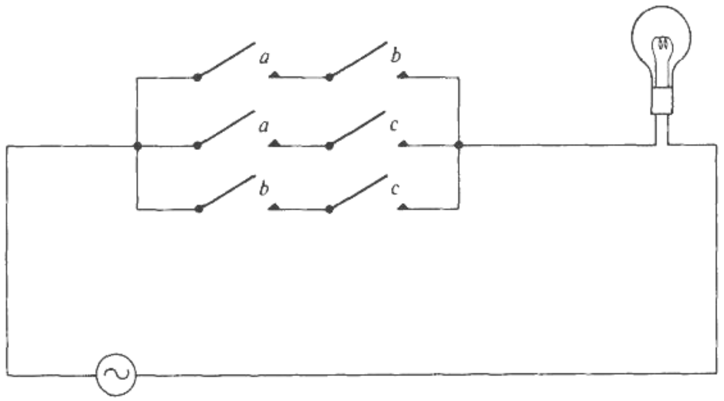


Figura 12.20

paralelo y serie de éstos, puede describirse mediante una expresión algebraica en nuestro sistema algebraico. Por ejemplo, la red de interruptoreo de la figura 12.18 puede ser descrita mediante la expresión algebraica:

$$a \textbf{ SER } ((b \textbf{ SER } c) \textbf{ PAR } (d \textbf{ SER } (e \textbf{ PAR } f)))$$

Nos damos cuenta de inmediato de que nuestro sistema algebraico es isomorfo al sistema algebraico $(\{0, 1\}, \wedge, \vee)$el cual se obtiene a partir de un álgebra booleana bivaluada, al eliminar la operación de negación. De aquí, tenemos primero que los resultados en álgebras booleanas pueden ser aplicados directamente al estudio del comportamiento de redes de interruptores. Además, vemos la posibilidad de utilizar interruptores para construir circuitos que realicen proposiciones que no incluyan la operación de negación. Por ejemplo, la figura 12.20 muestra un circuito el cual es un "tomador de la mayoría de votos", correspondiente a la proposición compuesta

$$(a \wedge b) \vee (a \wedge c) \vee (b \wedge c)$$

Esto es, la lámpara de la figura 12.20 se encenderá si y sólo si dos o más de los tres interruptores *a, b, c* están cerrados.

En la práctica, un interruptor de encendido-apagado se remplaza por un *relevador*. Un relevador es un dispositivo electromecánico mostrado esquemáticamente en la figura 12.2 la. Cuando el interruptor de encendido-apagado está cerrado, el solenoide electromagnético es energizado, lo cual, a su vez, cierra los contactos. En consecuencia, habrá un circuito cerrado entre las terminales 1 y 2. Cuando el interruptor de encendido-apagado está abierto, el solenoide electromagnético no es energizado y los contactos permanecerán abiertos. Por tanto, habrá un circuito abierto entre las terminales 1 y 2. Usaremos los mismos nombres simbólicos para referirnos tanto al interruptor de encendido-apagado que controla el solenoide electromagnético como a los contactos que controla el solenoide. En este momento, uno podría decir que un relevador funciona exactamente en la misma manera que lo hace un simple interruptor de encendido-apagado, y podemos preguntarnos por qué ir a través del paso "indirecto" de usar un interruptor de encendido-apagado para controlar los contactos. Una razón es que un solenoide electromagnético puede controlar más de un par de contactos. Por tanto, al cerrar el interruptor de encendido-apagado cerraremos más de un par de contactos, como muestra la figura 12.12*b*. Una aplicación simple es usar un interruptor

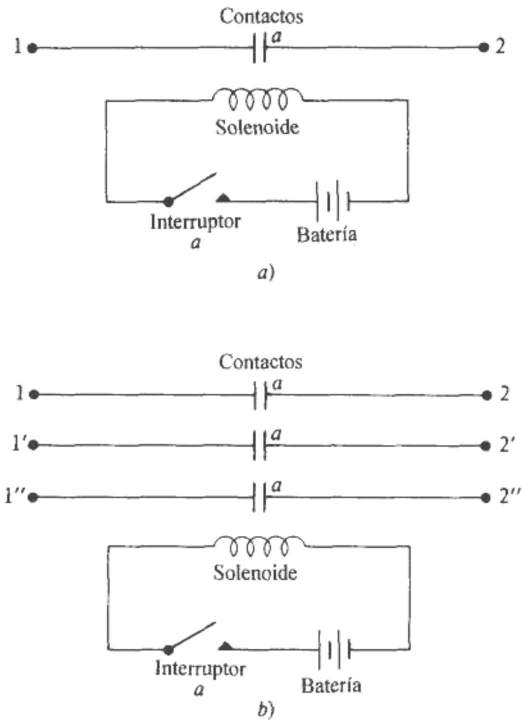


Figura 12.21

para encender varios dispositivos eléctricos simultáneamente, como un sistema estereofónico y varias lámparas. En segundo lugar tenemos que un relevador puede controlar dos tipos de contactos. Lo que acabamos de describir es conocido como contactos *abiertos normalmente,* lo que significa que éstos están abiertos cuando el interruptor controlador de encendido-apagado está abierto, y cerrados cuando el interruptor de encendido-apagado está cerrado. Como opción, podríamos tener los conocidos como contactos *cerrados normalmente.* Éstos están cerrados cuando el interruptor controlador de encendido-apagado está abierto, y abiertos cuando dicho interruptor está cerrado. Un par de contactos cerrados normalmente se muestra esquemáticamente en la figura 12.22, donde usamos *a* para denotar al interruptor de encendido-apagado y $\bar{a}$ para denotar el par de contactos cerrados normalmente controlados por *a.* Por ejemplo, uno puede utilizar un solo interruptor para encender un sistema estereofónico, y al mismo tiempo, apagar varias lámparas, o inversamente, como se muestra en la figura 12.23.

Podemos conectar contactos relevadores en paralelo y en serie exactamente de la misma manera en que son interconectados los interruptores de encendido-apagado. Por ejemplo, la

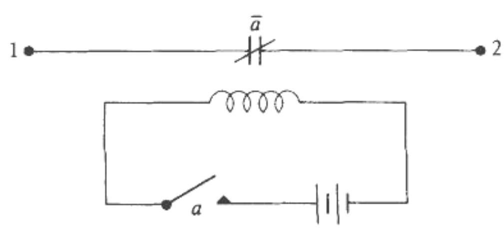


Figura 12.22

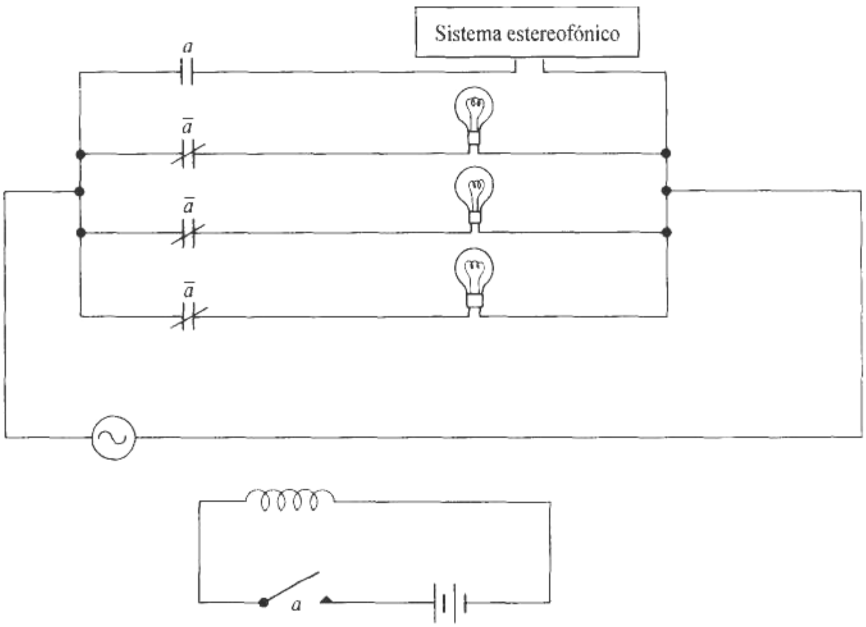


Figura 12.23

figura 12.24 muestra un circuito el cual es un tomador de votos para dos-de-tres. Señalemos que la lámpara se encenderá si y sólo si se cierran exactamente dos de tres de los interruptores de encendido-apagado *a, b, c.*

Observamos de inmediato que los circuitos constituidos de contactos relevadores

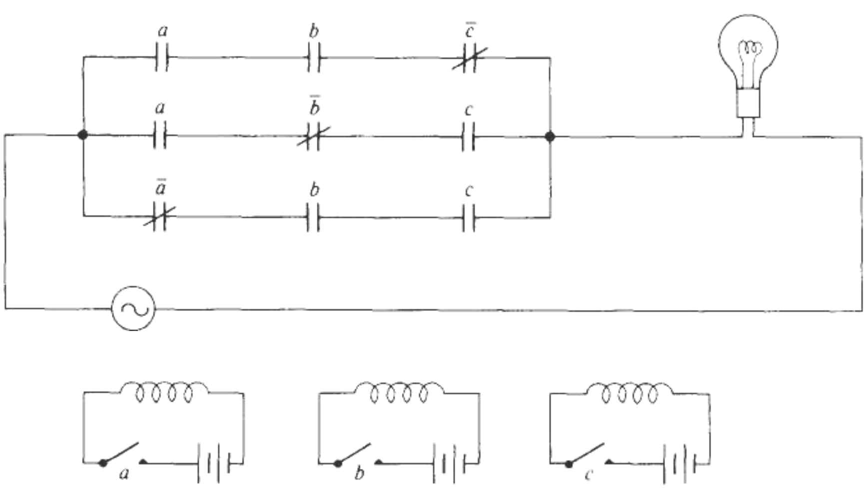


Figura 12.24

conectados en paralelo y en serie pueden ser descritos mediante expresiones algebraicas en el sistema algebraico ({abierto, cerrado}, **SER**, **PAR**, ¯), donde ¯ es una operación unaria correspondiente al uso de un par de contactos cerrados normalmente. El sistema algebraico ({abierto, cerrado}, **SER**, **PAR**, ¯) se conoce como el *álgebra de interruptores.* Sobra decir que el álgebra de interruptores es un álgebra booleana bivaluada, como ha anticipado el lector. En consecuencia todos los resultados de álgebras booleanas pueden ser aplicados al análisis y síntesis de redes de relevadores. Concluimos nuestro análisis con un ejemplo ilustrativo:

Ejemplo 12.4

Una lámpara es controlada mediante el circuito de interruptores mostrado en la figura 12.25a. El circuito puede ser descrito mediante la expresión algebraica:

$$(a\ \mathbf{PAR}\ \overline{b}\ \mathbf{PAR}\ d)\ \mathbf{SER}\ (a\ \mathbf{PAR}\ b\ \mathbf{PAR}\ \overline{c})\ \mathbf{SER}\ (a\ \mathbf{PAR}\ c\ \mathbf{PAR}\ d)$$

en el álgebra de interruptores. Esta expresión corresponde a la expresión

$$(a \vee \overline{b} \vee d) \wedge (a \vee b \vee \overline{c}) \wedge (a \vee c \vee d)$$

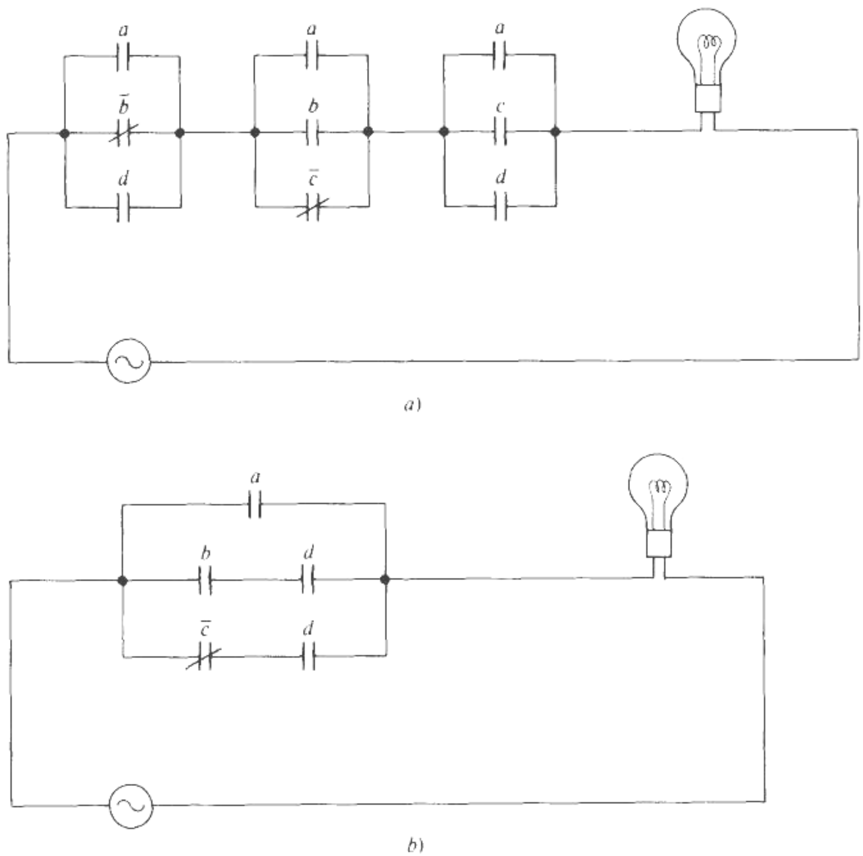


Figura 12.25

en el álgebra booleana bivaluada $(\{0, 1\}, \vee, \wedge, ^{-})$. Además,

$$\begin{aligned}
&(a \vee \bar{b} \vee d) \wedge (a \vee b \vee \bar{c}) \wedge (a \vee c \vee d) \\
&\qquad = [a \vee ((\bar{b} \vee d) \wedge (b \vee \bar{c}))] \wedge (a \vee c \vee d) \\
&\qquad = a \vee [(\bar{b} \vee d) \wedge (b \vee \bar{c}) \wedge (c \vee d)] \\
&\qquad = a \vee [[(\bar{b} \wedge \bar{c}) \vee (b \wedge d) \vee (\bar{c} \wedge d)] \wedge (c \vee d)] \\
&\qquad = a \vee (\bar{b} \wedge \bar{c} \wedge d) \vee (b \wedge c \wedge d) \vee (b \wedge d) \vee (\bar{c} \wedge d) \\
&\qquad = a \vee (b \wedge d) \vee (\bar{c} \wedge d)
\end{aligned}$$

Puesto que la expresión

$$a \vee (b \wedge d) \vee (\bar{c} \wedge d)$$

corresponde a la expresión

$$a \textbf{ PAR } (b \textbf{ SER } d) \textbf{ PAR } (\bar{c} \textbf{ SER } d)$$

en el álgebra de interruptores, el circuito de la figura 12.256 realizará la misma función de control que el de la figura 12.25*a*.

❐

12.11 NOTAS Y REFERENCIAS

Algunas referencias generales útiles sobre teoría de lattices y álgebras booleanas son Abbott [1], Birkhoff [2], Hohn [5], Rutherford [9] y Szász [10]. Para un análisis más extenso sobre el tópico de teoría de interruptores y diseño lógico véase, por ejemplo, Caldwell [3], Hill y Peterson [4] y Kohavi [7].

1. Abbott, J. C: *Sets, Lattices, and Boolean Algebras,* Allyn and Bacon, Boston, 1969.
2. Birkhoff, G.: *Lattice Theory,* 3ª ed., Am. Math. Soc. Coll. Publ., Providence, R. I., 1967.
3. Caldwell, S. H.: *Switching Circuits and Logical Design,* John Wiley & Sons, Nueva York, 1958.
4. Hill, F. J. y G. R. Peterson: *Introduction to Switching Theory and Logical Design,* 3ª ed., John Wiley & Sons, Nueva York, 1981.
5. Hohn, F. E.: *Applied Boolean Algebra,* Macmillan Company, Nueva York, 1960.
6. Guillemin, E. A.: *Introductory Circuit Theory,* John Wiley & Sons, Nueva York, 1953.
7. Kohavi, Z.: *Switching and Finite Automata Theory,* 2ª ed., McGraw-Hill Book Company, Nueva York, 1978.
8. Liu, C. L.: *Introduction to Combinatorial Mathematics,* McGraw-Hill Book Company, Nueva York, 1968.
9. Rutherford, D. E.: *Introduction to Lattice Theory,* Oliver and Boyd, Londres, 1965.
10. Szász, G.: *Introduction to Lattice Theory,* Academic Press, Nueva York, 1963.

PROBLEMAS

12.1 Sean a y b dos elementos de un lattice $(A, \leq)$. Demuestre que $a \wedge b = b$ si y sólo si $a \vee b = a$.

12.2 Sean a, b, c elementos de un lattice $(A, \leq)$. Demuestre que, si $a \leq b$, entonces $a \vee (b \wedge c) \leq b \wedge (a \vee c)$.

12.3 Sean a, b, c elementos en un lattice $(A, \leq)$. Demuestre que

$$a \vee (b \wedge c) \leq (a \vee b) \wedge (a \vee c)$$
$$(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee c)$$

12.4 Sean a y b dos elementos en un lattice $(A, \leq)$. Demuestre que $a \wedge b < a$ y $a \wedge b < b$ si y sólo si a y b son incomparables.

12.5 Sea $(A, \vee, \wedge)$ un sistema algebraico, donde $\vee$ y $\wedge$ son operaciones binarias que satisfacen la ley de absorción. Demuestre que $\vee$ y $\wedge$ también satisfacen la ley de idempotencia.

12.6 Estudiamos en este problema la posibilidad de definir un lattice mediante un sistema algebraico con dos operaciones binarias. Sea $(A, \vee, \wedge)$ un sistema algebraico, donde $\vee$ y $\wedge$ son operaciones binarias que satisfacen las leyes conmutativa, asociativa y de absorción.

a) Definimos una relación binaria $\leq$ sobre A tal que para todo a y b en A, $a \leq b$ si y sólo si $a \wedge b = a$. Demuestre que $\leq$ es una relación de orden parcial.

b) Demuestre que $a \vee b$ es la cota superior mínima de a y b en $(A, \leq)$. Demuestre que es $a \wedge b$ la cota inferior máxima de a y b en $(A, \leq)$.

12.7 Demuestre que un lattice es distributivo si y sólo si para a, b, c elementos cualesquiera en el lattice se cumple que

12.8 Sea $(A, \leq)$ un lattice distributivo. Demuestre que, si

$$a \wedge x = a \wedge y \quad \text{y} \quad a \vee x = a \vee y$$

para algún a, entonces $x = y$.

12.9 Demuestre que un lattice $(A, \leq)$ es distributivo si y sólo si para a, b, c elementos cualesquiera *en A,*

Sugerencia: para demostrar que $(A, \leq)$ es distributivo, considere los elementos a, $b \vee c$, y $(a \vee b) \wedge (a \vee c)$.

12.10 Un lattice $(A, \leq)$ recibe el nombre de *lattice modular* si para a, b, c elementos cualesquiera en A, donde $a \leq c$,

$$a \vee (b \wedge c) = (a \vee b) \wedge c$$

Demuestre que un lattice es modular si y sólo si la siguiente condición es válida:

$$a \vee (b \wedge (a \vee c)) = (a \vee b) \wedge (a \vee c)$$

12.11 Demuestre que para a, b, c elementos cualesquiera en un lattice modular

$$(a \vee b) \wedge c = b \wedge c$$

implica que

$$(c \vee b) \wedge a = b \vee a$$

12.12 Sea $(A, \leq)$ un lattice. Diremos que un subconjunto *I de A* es un *ideal* si se satisfacen las siguientes condiciones:

1. Para todo a y b en I, $a \vee b$ está en I.
2. Para todo a en I y para todo x en A, $a \wedge x$ está en I.

a) Demuestre que la condición 2 puede ser remplazada por:

2'. Si a está en I y $x \leq a$, entonces x está en I.

b) Sea I un subconjunto de A. Demuestre que I es un ideal si y sólo si para todo a y b en A:

1". Si a y b están en I, entonces a y b está en I.

2". Si $a \vee b$ está en I, entonces tanto a como b están en I.

12.13 Sea $(A, \leq)$ un lattice distributivo. Sean a y b dos elementos arbitrarios en A. Si $I(a, b)$ denota al conjunto

$$\{x | x \in A, a \wedge x = b \wedge x\}$$

Demuestre que $I(a, b)$ es un ideal (véase el problema 12.12 para la definición de un ideal).

12.14 Demuestre que en un lattice con una cota inferior universal 0 y una cota superior universal 1,0 es el *único* complemento de 1, y 1 es el *único* complemento de 0.

12.15 Demuestre que

$$a \vee (\bar{a} \wedge b) = a \vee b$$
$$a \wedge (\bar{a} \vee b) = a \wedge b$$

en un álgebra booleana.

12.16 Sea $(A, \vee, \wedge, \bar{\ })$ un álgebra booleana. Demuestre que $(A, \oplus)$ es un grupo conmutativo, donde $\oplus$ está definida como

$$a \oplus b = (a \wedge \bar{b}) \vee (\bar{a} \wedge b)$$

12.17 Sea $(A, \star)$ el sistema algebraico definido en el problema 11.13 [incluso con la condición de la parte *f*)] con la propiedad adicional de que para *a, b, c* en *A*,

$$a \star (b \star c) = b \star (a \star c)$$

decimos que $(A, \star$ es un *álgebra de implicación.* Definimos una relación binaria $\leq$ sobre *A* tal que $a \leq b$ si y sólo si $a \star b = e$.

a) Demuestre que $\leq$ ss una relación de orden parcial.
b) Demuestre que $a \leq b$ si y sólo si $b = x \star a$ para algún *x en A.*
c) Demuestre que para *a* y *b* en *A,* $(a \star b) \star b$ es una cota superior mínima de *a* y *b.*

12.18 Sea $(A, \vee, \wedge, \bar{\ })$ un álgebra booleana. Sea $\star$ una operación binaria definida sobre *A* tal que

$$a \star b = \bar{a} \vee b$$

Demuestre que $(A, \star)$ es un álgebra de implicación según la definición del problema 12.17.

12.19 Sea

$$E(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (\bar{x}_2 \wedge x_3)$$

una expresión booleana sobre el álgebra booleana bivaluada. Exprese $E(x_1, x_2, x_3)$ tanto en su forma normal disyuntiva como en su forma normal conjuntiva.

12.20 Sea

$$E(x_1, x_2, x_3) = \overline{(\overline{x_1 \vee x_2}) \vee (\bar{x}_1 \wedge x_3)}$$

una expresión booleana sobre el álgebra booleana bivaluada. Escriba $E(x_1, x_2, x_3)$ tanto en su forma normal disyuntiva como en su forma normal conjuntiva.

12.21 Sea

$$E(x_1, x_2, x_3, x_4) = (x_1 \wedge x_2 \wedge \bar{x}_3) \vee (x_1 \wedge \bar{x}_2 \wedge x_4) \wedge (x_2 \wedge \bar{x}_3 \wedge \bar{x}_4)$$

una expresión booleana sobre el álgebra booleana bivaluada. Exprese $E(x_1, x_2, x_3, x_4)$ tanto en su forma normal disyuntiva como en su forma normal conjuntiva.

12.22 *a)* Exprese la función de la figura 12P.1 en forma normal disyuntiva.
b) Exprese la función de la figura 12P. 1 en forma normal conjuntiva.

12.23 Simplifique las siguientes expresiones booleanas:
a) $(a \wedge b) \vee (a \wedge b \wedge c) \vee (b \wedge c)$
b) $a \wedge b) \vee (a \wedge \bar{b} \wedge c) \vee (b \wedge c)$

	f
(0, 0, 0)	1
(0, 0, 1)	0
(0, 1, 0)	1
(0, 1, 1)	0
(1, 0, 0)	0
(1, 0, 1)	1
(1, 1, 0)	0
(1, 1, 1)	1

Figura 12P.1

c) $a \wedge b) \vee (\bar{a} \wedge b \wedge \bar{c}) \vee (b \wedge c)$
d) $((a \wedge \bar{b}) \vee c) \wedge (a \vee \bar{b}) \wedge c$

12.24 En la Sección 12.7 definimos las nociones de forma normal disyuntiva y forma normal conjuntiva para expresiones booleanas sobre el álgebra booleana bi valuada. Veremos en este problema cómo estas nociones pueden extenderse a álgebras booleanas en general. Sea $E(x_1, x_2, \ldots, x_n)$ una expresión booleana sobre un álgebra booleana $(A, \vee, \wedge, \bar{\ })$. Usaremos la notación $E(x_i = a)$ para denotar a la expresión booleana obtenida a partir de $E(x_1, x_2, \ldots, x_n)$ mediante la sustitución de x_i por *a*, donde *a* es un elemento de *A*.

a) Para todo x_i demuestre que

$$E(x_1, x_2, \ldots, x_n) = (\bar{x}_i \wedge E(x_i = 0)) \vee (x_i \wedge E(x_i = 1))$$

Sugerencia: la igualdad puede demostrarse mediante inducción sobre la longitud de la expresión $E(x_1, x_2, \ldots, x_n)$, donde la longitud de una expresión se define como el número total de apariciones de elementos de *A*, nombres de variables, las operaciones $\vee, \wedge, \bar{\ }$ en la expresión tomando en cuenta las repeticiones. Así, expresiones de longitud uno son los elementos de *A* y las variables $x_1, x_2, \ldots, x_n$.

b) Emplee el resultado del inciso *a)* para demostrar que cualquier expresión booleana puede escribirse como una adición de expresiones de la forma

$$c_{\delta_1\delta_2\cdots\delta_n} \wedge \tilde{x}_1 \wedge \tilde{x}_2 \wedge \cdots \wedge \tilde{x}_n$$

donde $c_{\delta_1\delta_2\cdots\delta_n}$ es un elemento de *A*, y $\tilde{x}_i$ denota a x_i o bien $\bar{x}_i$. Tal forma de escribir una expresión booleana se conoce como una *forma normal disyuntiva.*

c) Sea

$$E(x_1, x_2) = (2 \wedge x_1) \wedge (x_1 \vee \bar{x}_2)$$

una expresión booleana sobre el álgebra booleana $(\{0, 1, 2, 3\}, \vee, \wedge, \bar{\ })$. Reescriba $E(x_1, x_2)$ en forma normal disyuntiva.

d) Sea *f* una función de A^n hacia *A* para el álgebra booleana $(A, \vee, \wedge, \bar{\ })$. Si *I* es una función booleana, ¿cómo podemos determinar una expresión booleana que especifique la función *f*?

Sugerencia: considere la forma normal disyuntiva de la expresión.

e) Demuestre que la función de la figura 12.8 no es una función booleana.

f) Para todo x_i, demuestre que

$$E(x_1, x_2, \ldots, x_n) = (x_i \vee E(x_i = 0)) \wedge (\bar{x}_i \vee E(x_i = 1))$$

g) Emplee el resultado del inciso/) para demostrar que cualquier expresión booleana puede escribirse como un producto de expresiones de la forma

$$d_{\delta_1\delta_2\cdots\delta_n} \vee \tilde{x}_1 \vee \tilde{x}_2 \vee \cdots \vee \tilde{x}_n$$

donde $d_{\delta_1\delta_2\cdots\delta_n}$ es un elemento de *A*, y denota $\tilde{x}_i$ a *x*, o bien $\bar{x}_i$. Tal forma de escribir una expresión booleana se conoce como una forma normal conjuntiva.

12.25 Un estudiante recibirá una calificación de aprobado en un curso si reúne una o más de las siguientes condiciones:

1. Obtuvo una calificación de B o mayor en el examen de mitad de periodo y una A en el examen final.
2. Obtuvo una calificación de B o mayor tanto en el examen de mitad de periodo como en el examen final y no falló en entregar alguna tarea.
3. Obtuvo una calificación de B o mayor en al menos uno de los exámenes de mitad de periodo y final, y tiene una beca de atletismo.
4. Su padre es el instructor del curso.

Utilice las siguientes proposiciones como variables para escribir una expresión booleana que describa el conjunto de condiciones indicadas arriba.

a: Obtener una B en el examen de mitad de periodo
b: Obtener una A en el examen de mitad de periodo
c: Obtener una B en el examen final
d: Obtener una A en el examen final
e: Fallar en entregar alguna tarea
f: Tener una beca de atletismo
g: Ser el hijo del instructor

Simplifique la expresión y obtenga así un conjunto más simple de condiciones.

12.26 La siguiente tabla muestra el contenido de cinco cajas de herramienta etiquetadas como *a, b, c, d* y *e:*

	Desarmador	Llave de tuercas	Tenazas	Martillo	Sierra
a	x	x			
b		x	x	x	
c	x		x		
d		x			x
e			x	x	

Escriba una expresión booleana para mostrar cuántas selecciones de cajas de herramientas se pueden hacer de manera que cada selección contenga al menos una herramienta de cada tipo.

12.27 Se va a seleccionar un comité a partir de los cinco candidatos *a, b, c, d, e.* La selección debe satisfacer todas las condiciones siguientes:

1. *a o b* deben ser incluidos, pero no ambos.
2. *c* o *e* o ambos deben ser incluidos.
3. Si *d* es incluido, entonces *b* debe ser incluido.
4. Ya sea que tanto *a* como c sean incluidos, o bien ninguno es incluido.
5. Si *e* es incluido, entonces *c* y *d* deben ser incluidos.

¿Cómo deberá hacerse la selección?

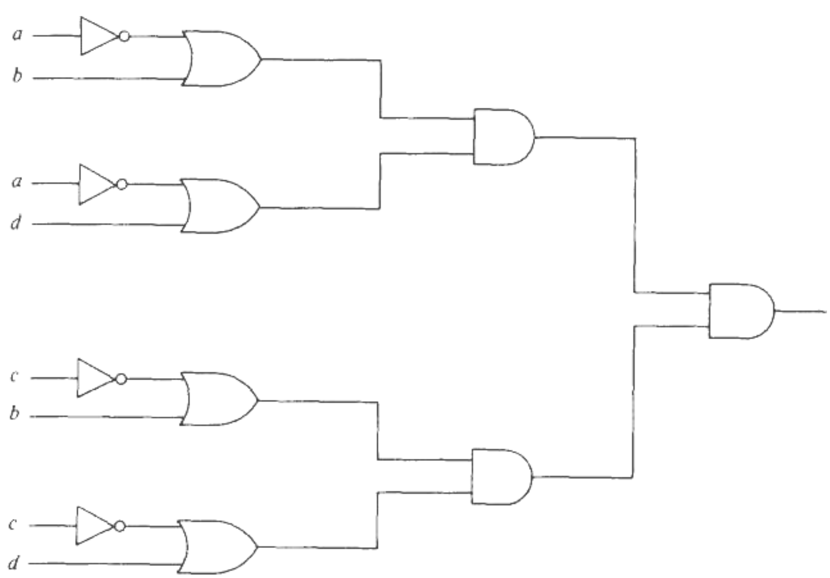


Figura 12P.2

12.28 Demuestre cómo el método de álgebras booleanas puede ser aplicado para determinar conjuntos dominantes minimales de un grafo. Ilustre el procedimiento mediante la determinación de los conjuntos dominantes minimales del grafo de la figura 5P.2 (véase el problema 5.11 para la definición de un conjunto dominante de un grafo).
Sugerencia: para cada vértice, el vértice mismo o bien alguno de los vértices adyacentes a éste, debe ser incluido en el conjunto dominante.

12.29 Simplifique el circuito de la figura 12P.2 (existe un circuito que utiliza solamente tres compuertas-OR y compuertas-AND).

12.30 *a)* Muestre cómo se puede remplazar una compuerta-OR mediante la adecuada interconexión de compuertas-AND y compuertas-NOT.
b) Demuestre cómo se puede remplazar una compuerta-AND mediante la adecuada interconexión de compuertas-OR y compuertas-NOT.

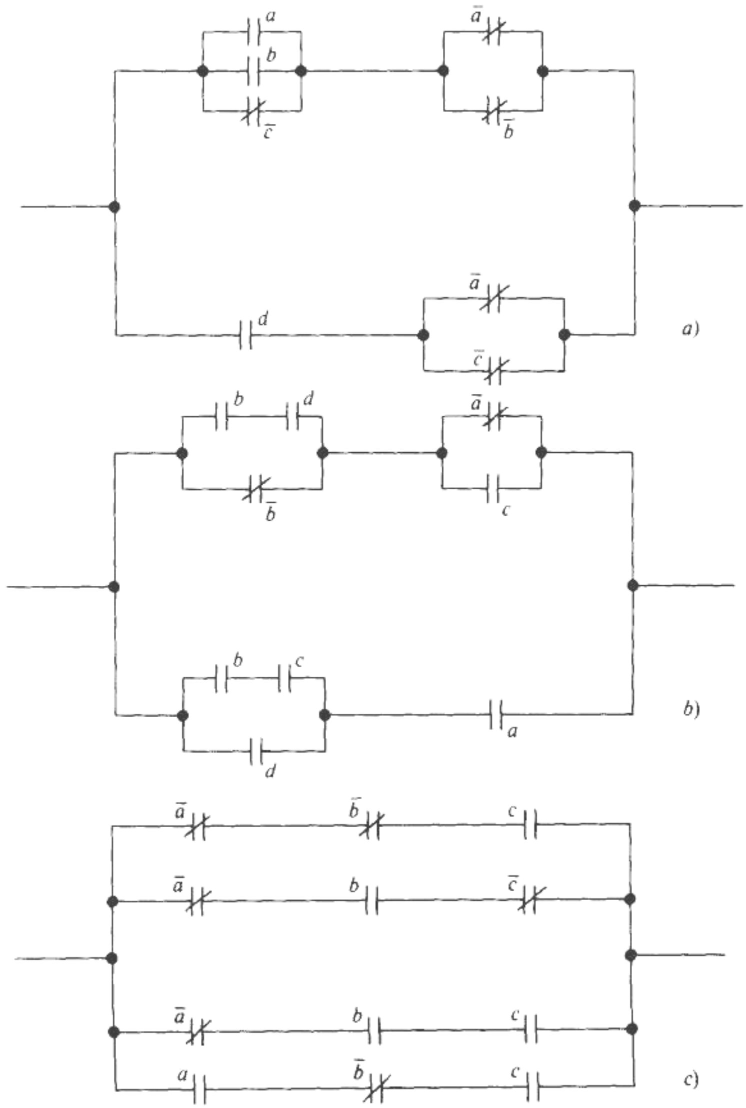


Figura 12P.3

c) Una compuerta-NOR es un circuito que tiene dos entradas y una salida de tal manera que la salida es igual a $\overline{x \vee y}$ para las entradas *x* y *y*. Demuestre que podemos realizar cualquier expresión booleana solamente con el empleo de compuertas-NOR, mostrando cómo se pueden remplazar las compuertas-OR, compuertas-AND, y compuertas-NOT mediante interconexiones adecuadas de compuertas-NOR.

d) Una compuerta-NAND es un circuito que tiene dos entradas y una salida de manera que la salida es igual a $\overline{x \wedge y}$ para las entradas *x* y *y*. Demuestre que podemos realizar cualquier expresión booleana solamente con el empleo de compuertas-NAND, mostrando cómo se pueden remplazar las compuertas-OR, compuertas-AND y compuertas-NOT mediante inter conexiones adecuadas de compuertas-NAND.

12.31 *a)* Simplifique los circuitos relevadores de la figura 12P.3.

b) Construya redes digitales con compuertas-OR, compuertas-AND y compuertas-NOT para efectuar las expresiones booleanas realizadas mediante las redes relevadoras de la figura 12P.3.

ÍNDICE

Segunda Edición

Elementos de Matemáticas Discretas

Esta obra tiene como finalidad que el estudiante aprenda las bases de las ***Matemáticas Discretas***. El autor dedica atención especial a un objetivo: que el lector descubra las aplicaciones posibles de esta ciencia. El libro incluye gran cantidad de ejercicios y ejemplos junto con las definiciones y desarrollos teóricos. Los temas que presenta incluyen las nociones para el desarrollo de tópicos tales como:

- Análisis combinatorio
- Álgebra moderna aplicada
- Teoría de compiladores
- Sistemas y estructuras de datos
- Teoría de autómatas
- Lenguajes formales
- Máquinas de estado finito

Este libro expone en forma realmente amena las bases de las matemáticas discretas que han sido el fundamento de la tecnología moderna.

9 789701 007433

ISBN: 970-10-0743-3

Mc Graw Hill