

TÉCNICAS DE SIMULACIÓN EN COMPUTADORAS "

Autor: Naylor, Thomas

CAPITULO 4

Generación de valores de las variables estocásticas empleadas en simulación

INTRODUCCION

Cuando se establecen las bases racionales subyacentes al empleo de los métodos existentes para generar valores de variables estocásticas en una computadora digital, se parte de dos problemas un tanto divergentes. Estos dos problemas de tipo distinto se pueden clasificar convenientemente como determinísticos, es decir, no probabilísticos o bien, como estocásticos. Recientemente se ha popularizado el término Monte Carlo como sinónimo para el concepto *simulación de procesos estocásticos*. Sin embargo, conviene anotar que en el pasado, este término se aplicó tan sólo al emplear los métodos de simulación estocástica para la resolución de problemas estrictamente determinísticos.

En un principio, los métodos de simulación estocástica fueron aplicados por los matemáticos y los científicos relacionados con las áreas de la Física, para resolver ciertos problemas determinísticos que se podían expresar mediante ecuaciones matemáticas para las cuales sus soluciones no resultaban fáciles de obtener, utilizando los criterios convencionales de los métodos numéricos o analíticos. Cabe considerar el hecho de que para cierto número de problemas matemáticos de importancia reconocida, existe la posibilidad de que, una vez encontrado un proceso estocástico cuya distribución de probabilidad o cuyos parámetros satisfagan las propiedades matemáticas que se requieran, quedan resueltas las ecuaciones que caracterizan a estos problemas. Aún más, desde un punto de vista computístico pudiera resultar más eficiente construir tal tipo de procesos a la vez que generar su estadística, empleando la computadora en lugar de seguir los métodos convencionales. Entre los problemas matemáticos determinísticos para los que se ha encontrado que la simulación estocástica resulta útil en la obtención de

donde e_i es una constante y Π denota al producto $p_1^{e_1} \times p_2^{e_2} \times p_3^{e_3} \dots$. La prueba de esto se debe a Euclides [2, p. 21].

Teorema 5. Si $(a, m) = 1$, entonces $a^{\varphi(m)} \equiv 1 \pmod{m}$, de lo cual se sigue que:

(1) El mayor orden posible de a es $h = \varphi(m)$ cuando a es una raíz primitiva de m .

(2) Para $n < m$ tales que $(m, n) = 1$, $na^h \equiv n \pmod{m}$, donde $h = \varphi(m)$. La prueba de esto se atribuye a Euler [21, p. 273] y se obtiene de los teoremas 2 y 3.

Teorema 6. Para todas las potencias de un número primo $p > 2$ existen las raíces primitivas, i.e. existe un número tal que $(a, p^e) = 1$ y $a^{\varphi(p^e)} \equiv 1 \pmod{p^e}$ donde $h = \varphi(p^e)$. (Véase [21, p. 285].)

Teorema 7. Si $m = \Pi p_i^{e_i}$, entonces $\varphi(m) = \Pi (p_i - 1) p_i^{e_i - 1}$. La demostración se debe a Euler [21, p. 113].

Teorema 8. Si $m = p^e$ y p es un primo impar, entonces $h = \lambda(m) = (p - 1) p^{e-1} = \varphi(m)$ para valores de a que son raíces primitivas de m . Corolario: Si $p = 2$, i.e., $h = \lambda(m) = 2^{e-2}$ para $e > 2$, entonces $\lambda(m) \neq \varphi(m)$. La prueba se debe a Euler [21, pp. 283-290].

Teorema 9. Si $m = \Pi p_i^{e_i}$ para $i = 1, 2, \dots, s$, entonces:

(1) $\lambda(m) = \text{m.c.m}[\lambda(p_1^{e_1}), \lambda(p_2^{e_2}), \dots, \lambda(p_s^{e_s})]$.

(2) Existen valores de a cuyo orden es igual a (esto es, pertenecen conjuntamente a) cada $\lambda(p_i^{e_i})$. La demostración está en [21, p. 293] y se sigue del teorema chino del residuo debido a Sun-Tse [21, p. 246].

Colorario: Si $p_1 = 2$, entonces $\lambda(m) = \text{m.c.m}[\lambda(2^{e_1}), \varphi(p_2^{e_2}), \varphi(p_3^{e_3}), \dots]$.

Teorema 10. El menor entero positivo a tal que $(a^h - 1)/(a - 1) \equiv 0 \pmod{m}$ es $h = m$, si (1) $a \equiv 1 \pmod{p}$ si p es un factor primo de m y (2) $a \equiv 1 \pmod{4}$ si 4 es un factor de m . La prueba se debe a Hull y Dobell [16, pp. 233-235].

REFERENCIAS Y BIBLIOGRAFIA

1. Allard, J. L., Dobell, A. R. y Hull, T. E. "Mixed Congruential Random Number Generators for Decimal Machines", *Journal of the Association for Computing Machinery*, X, No. 2 (1963), 131-141.
2. Birkhoff, G., y MacLane, S. *A Survey of Modern Algebra*. Nueva York: The Macmillan Company, 1953.
3. Coveyou, R. R., "Serial Correlation in the Generation of Pseudo-Random Numbers", *Journal of the Association for Computing Machinery*, VII (1960), 72-74.
4. Duparc, H., Lekkerkerker, C. G. y Peremans, W. "Reduced Sequences of Integers and Pseudo-Random Numbers", *Mathematische Centrum Report ZW 1953-002*, (1953).
5. Fisher, R. A., y Yates, F. *Statistical Tables for Biological Agricultural and Medical Research*. Londres: Oliver and Boyd, 1953.

6. Forsythe, G. E. "Generation and Testing of Random Digits at the National Bureau of Standards, Los Angeles", en *Monte Carlo Method*. National Bureau of Standards Applied Mathematics Series No. 12. Washington, D. C., 1951.
7. Freund, J. E. *Mathematical Statistics*. Englewood Cliffs: Prentice-Hall, 1962.
8. Golenko, D. K., y Smiriagin, V. A. "A Source of Random Numbers Which Are Equidistributed in $[0, 1]$ ", *Publications Math. Inst. Hungarian Acad. Sci.* 5, Series A. Fasc. 3, en ruso, con resumen en inglés (1960), 241-253.
9. Good I. J. "The Serial Test for Sampling Numbers and Other Tests of Randomness" *Proc. Camb. Phil. Soc.*, XLIX (1953), 276-284.
10. Good, I. J. "On the Serial Test for Random Sequences", *Annals of Mathematical Statistics* XXVIII (1957), 262-264.
11. Green, B. F., *Digital Computers in Research*. Nueva York: McGraw-Hill Book Co., 1963.
12. Green, B. F., Smith J., y Klem L. "Empirical Tests of an Additive Random Number Generator", *Journal of the Association for Computing Machinery*, VI, No. 4 (1959), 527-537.
13. Greenberger, M. "An a Priori Determination of Serial Correlation in Computer Generated Random Numbers", *Mathematics of Computations*, XV (1961), 383-389.
14. Greenberger, M., "Method in Randomness", *Communications of the ACM*, VIII, No. 3 (1965), 177-179.
15. Hull, T. E., y Dobell, A. R. "Mixed Congruential Random Number Generators for Binary Machines", *Journal of the Association for Computing Machinery*, XI, No. 1 (1964), 31-40.
16. Hull, T. E., y Dobell, A. R. "Random Number Generators", *SIAM Review*, IV, No. 3 (julio, 1962), 230-254.
17. International Business Machines Corporation, "Random Number Generation and Testing", Reference Manual (C20-8011), Nueva York, 1959.
18. Lehmer, D. H. "Mathematical Methods in Large-Scale Computing Units", *Annals Computer Laboratory Harvard University*, XXVI (1951), 141-146.
19. MacLaren, M. D., y Marsaglia, G. "Uniform Random Number Generators", *Journal of the Association for Computing Machinery*, XII, No. 1 (1965), 83-89.
20. National Bureau of Standards. *Monte Carlo Method*. Applied Mathematics Series No. 12. Washington, D. C., 1951.
21. Ore, O. *Number Theory and Its History*. Nueva York: McGraw-Hill Book Co., 1948.
22. RAND Corporation. *A Million Random Digits with 100,000 Normal Deviates*. Glencoe, Ill.: The Free Press, 1955.
23. Rotenberg, A. "A New Pseudo-Random Number Generator", *Journal of the Association for Computing Machinery*, VII (1960), 75-77.
24. Stockmal, F. "Calculations with Pseudo-Random Numbers", *Journal of the Association for Computing Machinery*, XI, No. 1 (enero, 1964), 41-52.
25. Taussky, O., y Todd, J. "Generation and Testing of Pseudo-Random Numbers" en *Symposium on Monte Carlo Methods*, ed. Herbert A. Meyer. Nueva York: John Wiley and Sons, Inc., 1956.
26. Tocher, K. D. "The Application of Automatic Computers to Sampling Experiments", *Journal of the Royal Statistical Society*, B16 (1954), 39-61.
27. Uspensky, James V., y Heaslet, M. A. *Elementary Number Theory*. Nueva York: McGraw-Hill Book Co., 1939.
28. Wold, H. "Random Normal Deviates", *Tracts for Computers*, No. XXV. Londres: Cambridge University Press, 1955.

o alguna de sus variantes respectivas, proporcionan la base general para simular la mayoría de las distribuciones consideradas en este capítulo.

El método de la transformación inversa

Si deseamos generar los valores x_i de las variables aleatorias a partir de cierta estadística de población cuya función de densidad esté dada por $f(x)$, debemos en primer lugar obtener la función de distribución acumulativa $F(x)$. (Véase la figura 4-1.) Puesto que $F(x)$ se define sobre el rango de 0 a 1, podemos generar números aleatorios distribuidos uniformemente y además hacer $F(x) = r$. Resulta claro, entonces, cómo queda x determinada

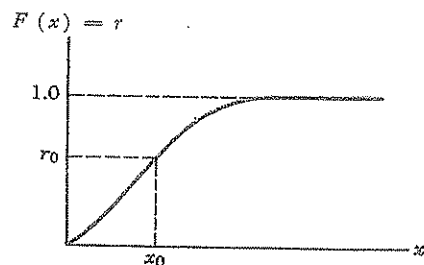


Figura 4-1. Una función de distribución acumulativa.

unívocamente por $r = F(x)$. Sigue, por lo tanto, que para cualquier valor particular de r , que generemos, por ejemplo r_0 , siempre es posible encontrar el valor de x ; en este caso x_0 , que corresponde a r_0 debido a la función inversa de F , si es conocida. Esto es,

$$x_0 = F^{-1}(r_0) \quad (4-2)$$

donde $F^{-1}(x)$ es la transformación inversa (o mapeo) de r sobre el intervalo unitario en el dominio de x . Si generamos números aleatorios uniformes correspondientes a una $F(x)$ dada, podemos resumir, matemáticamente, este método como sigue:

$$r = F(x) = \int_{-\infty}^x f(t) dt \quad (4-3)$$

entonces

$$P\{X \leq x\} = F(x) = P[r \leq F(x)] = P[F^{-1}(r) \leq x], \quad (4-4)$$

y consecuentemente $F^{-1}(r)$ es una variable que tiene a $f(x)$ como función de densidad de probabilidad. Este criterio equivale a resolver la ecuación (4-3) en x , en términos de r ; el procedimiento se ilustra con los siguientes ejemplos:

Ejemplo 1. Genérense los valores x de variables aleatorias con una función de densidad $f(x) = 2x$, $0 \leq x \leq 1$. De la ecuación (4-3) resulta que

$$r = F(x) = \int_0^x 2t dt \quad 0 \leq x \leq 1 \quad (4-5)$$

$$= x^2.$$

Tomando ahora la transformación inversa $F^{-1}(r)$, esto es, resolviendo la ecuación (4-5) para x , obtendremos

$$x = F^{-1}(r) = \sqrt{r}, \quad 0 \leq r \leq 1. \quad (4-6)$$

Por lo tanto, los valores de x con una función de densidad $f(x) = 2x$ se pueden generar al determinar la raíz cuadrada de los números aleatorios r .

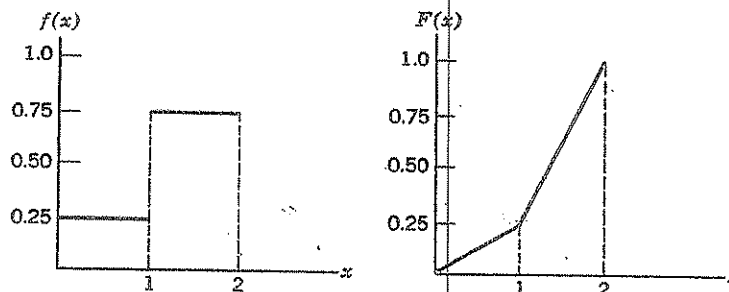


Figura 4-2. La función de densidad y la función de distribución acumulativa, para el ejemplo 2.

Ejemplo 2. Genérense los valores x de variables aleatorias con una función de densidad

$$f(x) = \frac{1}{4} \quad 0 \leq x \leq 1 \quad (4-7)$$

$$= \frac{3}{4} \quad 1 \leq x \leq 2$$

(La función de densidad y la función de distribución acumulativa se ilustran gráficamente en la figura 4-2.) A partir de la ecuación (4-3) sigue que

$$r = F(x) = \int_0^x \frac{1}{4} dt \quad 0 \leq x < 1 \quad (4-8)$$

$$= \frac{x}{4}$$

$$r = F(x) = \frac{1}{4} + \int_1^x \frac{3}{4} dt \quad 1 \leq x \leq 2 \quad (4-9)$$

$$= \frac{3}{4}x - \frac{1}{2}$$

Si tomamos la transformación inversa $F^{-1}(r)$ y resolvemos en x las ecuaciones (4-8) y (4-9), obtenemos

$$x = 4r \quad 0 \leq r < \frac{1}{4} \quad (4-10)$$

$$x = \frac{4}{3}r + \frac{2}{3} \quad \frac{1}{4} \leq r \leq 1. \quad (4-11)$$

Para generar un valor de x deberemos, en primer lugar, generar un valor de r ; cuando r sea menor que $1/4$, el valor de x estará determinado por la ecuación (4-10). Si r es mayor o igual a $1/4$, entonces x estará determinada por la ecuación (4-11). Para el caso de los intervalos múltiples sobre la escala caracterizada por r , podemos fácilmente generalizar este procedimiento a fin de generar los valores de las variables aleatorias que correspondan a una determinada distribución empírica.

Desafortunadamente, para muchas de las distribuciones de probabilidad, resulta imposible o extremadamente difícil expresar a x en términos de la transformación inversa $F^{-1}(r)$. Cuando es éste el caso, el único curso de que disponemos consiste en obtener una aproximación numérica para la transformación inversa $F^{-1}(r)$, o bien recurrir a alguno de los siguientes dos métodos.

El método de rechazo

Si $f(x)$ es una función acotada y x tiene además un rango finito, con $a \leq x \leq b$, entonces se puede utilizar la técnica de rechazos [29] para generar los valores de variables aleatorias. La aplicación de esta técnica requiere que se proceda de acuerdo con las siguientes etapas:

1. Normalizar el rango de f mediante un factor de escala c tal que

$$c \cdot f(x) \leq 1 \quad a \leq x \leq b. \quad (4-12)$$

2. Definir a x como una función lineal de r , o sea

$$x = a + (b - a)r. \quad (4-13)$$

3. Generar parejas de números aleatorios (r_1, r_2) .

4. Siempre que se encuentre una pareja de números aleatorios satisfagan la relación

$$r_2 \leq c \cdot f[a + (b - a)r_1]. \quad (4-14)$$

dicho par será aceptado y se utilizará a $x = a + (b - a)r_1$ como el valor generado de la variable aleatoria.

La teoría sobre la que se apoya este método se basa en el hecho ya conocido relativo a la probabilidad de que r sea menor o igual a $c \cdot f(x)$ es

$$P[r \leq c \cdot f(x)] = c \cdot f(x). \quad (4-15)$$

En consecuencia, si x se elige al azar dentro del rango (a, b) de acuerdo con la ecuación (4-13) y en el caso de que $r > c \cdot f(x)$ se rechaza, la función de densidad de probabilidad de los valores de x aceptados deberá ser igual a $f(x)$. Tocher [28, p. 25] ha demostrado que la esperanza matemática del número de intentos que se realizan, antes de encontrar una pareja exitosa, es igual a $1/c$. Esto implica que para ciertas funciones de densidad de probabilidad este método puede resultar sumamente ineficaz. En muchas de las técnicas generatrices que se describen en este capítulo se empleará este método, para lo cual se incluyen dos ejemplos a fin de aclarar ideas sobre él.

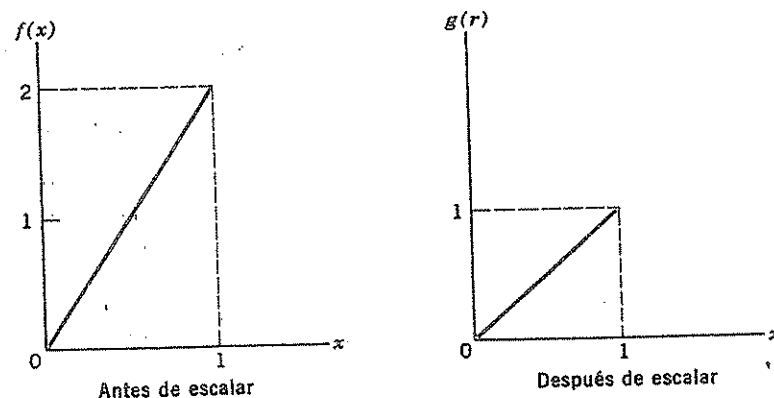


Figura 4-3. Un ejemplo de escalamiento.

Ejemplo 1. Utilice el método de rechazo para generar los valores x de las variables aleatorias, con una función de densidad $f(x) = 2x$ para las que $0 \leq x \leq 1$.

Puesto que x ha sido definida sobre el intervalo unitario, se tendrá que $x = r$; pero $f(r) = 2r$ está definida en el intervalo $0 \leq f(r) \leq 2$. En consecuencia, si se escala haciendo $g(r) = 1/2 f(r)$, se transformará a $f(r)$ al intervalo unitario, en cuyo caso $g(r) = r$. La figura 4-3 muestra la función de densidad $f(x) = 2x$, antes del proceso de escalamiento y después de él.

Para el ejemplo 1 el método de rechazo consta de cuatro pasos:

1. Generar r_1 y calcular $g(r_1)$.

2. Generar r_2 y compararlo con $g(r_1)$.
3. Si $r_2 \leq g(r_1)$, se acepta a r_1 tomándolo como una x de $f(x)$; Si $r_2 > g(r_1)$, se rechaza a r_1 y se vuelve a empezar por el paso 1.
4. Este proceso se repite hasta haber generado n valores de x . El método de rechazo se puede utilizar también, al igual que la técnica Monte Carlo, para evaluar integrales definidas.

Ejemplo 2. Use el método de rechazo para computar el área correspondiente al primer cuadrante de un círculo unitario cuyos ejes coordenados son r_1 y r_2 , respectivamente (figura 4-4). Este problema de integración numérica servirá para ilustrar el empleo del método Monte Carlo en la solución de un problema completamente determinístico. Cualquier pareja de números aleatorios (r_1, r_2) definidos en el intervalo unitario, corresponde a un punto contenido en el cuadrado unitario de la figura 4-4; obsérvese que los puntos que satisfacen la ecuación $r_1^2 + r_2^2 = 1$ se encuentran en la circunferencia. Sea $g(r_1) = \sqrt{1 - r_1^2}$; si para los números aleatorios generados (r_1^0, r_2^0) se tiene que $g(r_1^0) \geq r_2^0$, entonces (r_1^0, r_2^0) será un punto aleatorio bajo la curva. Pero si $g(r_1^0) < r_2^0$, entonces (r_1^0, r_2^0) será un punto que se encuentre sobre la curva. Al aceptar y contar el primer tipo de ocurrencias aleatorias y luego dividir esta cuenta entre el número total de parejas generadas, obtenemos un cociente que corresponde a la proporción del área del cuadrado unitario que se encuentra bajo la curva.

Este cociente se aproximará al valor de $\pi/4$, a medida que se incremente el número de pares de números aleatorios aceptados. Esta misma técnica se puede aplicar a la solución de integrales múltiples de funciones con más de una variable independiente.

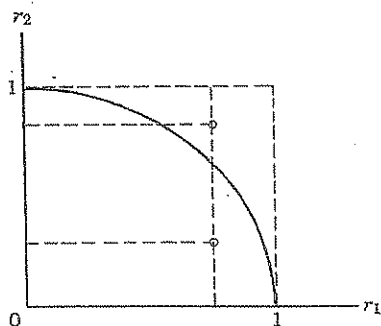


Figura 4-4. Integración numérica.

El método de composición

Otro método para generar valores de variables estocásticas utilizando computadoras es el llamado método de composición o método de mezclas [7,

13, 18, 19, 28]. En este método se expresa a $f(x)$ como una mezcla probabilística de las funciones de densidad $g_n(x)$ seleccionadas adecuadamente. En términos matemáticos tenemos

$$f(x) = \sum g_n(x) p_n \quad (4-16)$$

La guía para la selección de las $g_n(x)$ está dada sobre las consideraciones relativas a la bondad del ajuste y al objetivo de minimizar $\sum T_n p_n$, donde

T_n es el tiempo esperado de computación para generar valores de variables aleatorias a partir de $g_n(x)$.

En las partes subsecuentes de este capítulo, se proveerá al lector de un conjunto de (relativamente simples) técnicas específicas para simular valores de variables aleatorias, considerando algunas de las distribuciones de probabilidad mejor conocidas. En el caso de algunas distribuciones, se considerará más de un método alternativo. Intentamos desplazarnos desde las distribuciones de probabilidad específicas hasta los modelos estocásticos en general, con el fin de ampliar nuestro estudio sobre las técnicas de simulación. Para una revisión de los elementos de la teoría de probabilidades, remitiremos al lector a los textos siguientes [1, 8, 10, 23, 31].

En principio, se cubrirán separadamente las distribuciones de probabilidad continua y discreta. Primeramente se tratarán seis de las distribuciones continuas más comunes: la uniforme, exponencial, gamma, normal, normal multivariada y normal logarítmica. Para cada una de ellas se proporcionará la siguiente información: (1) una breve descripción relativa a la naturaleza y uso de la distribución; (2) las fórmulas para la función de densidad, la función de distribución acumulativa (si es que existe en forma explícita), el valor de la esperanza matemática y la variancia de la distribución; (3) los parámetros de la distribución, expresados en términos de los momentos de la distribución; (4) una explicación o en su defecto una derivación, de las técnicas más simples para generar los valores de las variables aleatorias de acuerdo con la distribución; (5) un diagrama de flujo y un programa en FORTRAN para generar los valores de las variables aleatorias mediante una computadora digital; (6) algunas técnicas alternativas para generar los mismos valores; (7) una lista útil de los valores de las variables aleatorias relacionadas o derivables de los valores encontrados en (el caso de que existan). Se respetará un formato semejante en el tratamiento de las cinco distribuciones discretas de probabilidad: la geométrica, de Pascal, la binomial, hipergeométrica y de Poisson. Se dedicarán secciones especiales para las distribuciones empíricas, los procesos de Markov y los valores de las variables aleatorias autocorrelacionados.

Aunque este capítulo está orientado a la utilización de las computadoras digitales en la simulación de distribuciones de probabilidad, la computadora no constituye, en forma alguna, un requisito previo para emplear las técnicas que se encuentran en este capítulo. Para mayor seguridad, en cualesquiera de los métodos que aquí se tratan, se pueden utilizar las técnicas de computación manual. Empero, si el número de las distribuciones de probabilidad que se van a simular es muy grande, y la cantidad de datos por simular es considerable, será imperioso el empleo de la computadora.

En este libro se seleccionó el sistema de programación FORTRAN debido a que es un lenguaje de computadora ampliamente utilizado, que se

asemeja mucho al lenguaje de las matemáticas y fue diseñado en principio para los procesos de computación tanto científicos como ingenieriles. Una de las ventajas principales de FORTRAN es la de proporcionar al analista un medio eficiente para escribir sus programas de computadora. Además, no requiere para su uso un período muy largo de instrucción, así como tampoco algún conocimiento detallado de la propia computadora. Aún más, los compiladores FORTRAN se encuentran en la actualidad disponibles para casi todas las computadoras en uso, ya sea en la industria, en las dependencias del Gobierno y en las universidades. En este libro, se encuentra un lenguaje FORTRAN que no necesariamente está diseñado para una computadora en particular y con muy pocas modificaciones se puede adaptar al de cualquier máquina que tenga un compilador FORTRAN. Debido a esto, las proposiciones FORTRAN que aparecen en este libro se han mantenido, deliberadamente, con una estructura muy simple y sin que requieran el empleo de los medios de entrada y salida. El lector que no esté familiarizado con el FORTRAN podrá consultar el respectivo manual, publicado por algunos fabricantes de computadoras, o alguno de los siguientes textos sobre FORTRAN [15, 22].

En este capítulo las proposiciones FORTRAN se presentarán como subrutinas (SUBROUTINE) y supone que existe un programa principal de simulación llamado MAIN, el cual llama a las apropiadas subrutinas mediante la proposición CALL. Cada subrutina, genera y envía al programa principal un solo valor de la variable aleatoria, a partir de una distribución de probabilidad para la cual fue programada. El programa principal deberá contener las instrucciones que permitan leer los valores de los parámetros que se requieran, el número de valores de las variables aleatorias que se deban generar y los criterios acerca de cómo manipular los resultados estadísticos.

A fin de evitar complicaciones notacionales al escribir subrutinas FORTRAN que contengan otras subrutinas, se establecerá que los números pseudoaleatorios están generados por una función propia del compilador y disponible en un programa de biblioteca previamente programado. Esta función se denota por RND y se supone programada en lenguaje de máquina de acuerdo con el criterio de alguno de los métodos propuestos en el capítulo 3. Otras funciones de biblioteca de las que se puede disponer también en forma de subrutinas son, la función logarítmica (de base e) denotada por LOG y la función exponencial (en base e), representada por EXP.

Ahora podemos enfocar nuestra atención a la tarea de desarrollar los métodos previamente mencionados, relativos a la generación de los valores de variables aleatorias a partir de las distribuciones de probabilidad.

DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

La distribución uniforme

Quizá la función de densidad de probabilidad más simple es aquella que se caracteriza por ser constante, en el intervalo (a, b) y cero fuera de él. Esta función de densidad define la distribución conocida como uniforme o rectangular. La distribución uniforme surge cuando se estudian las características de los errores por redondeo al registrar un conjunto de medidas sujetas a cierto nivel de precisión. Por ejemplo, si se registran medidas de peso con una aproximación determinada en gramos, puede suponerse que la diferencia en gramos entre el peso real y el peso registrado corresponde a un cierto número entre -0.5 y $+0.5$, y que el error se encuentra distribuido uniformemente en este intervalo. El valor más sobresaliente que puede tener la distribución uniforme respecto a las técnicas de simulación radica en su simplicidad y en el hecho de que tal distribución se puede emplear para simular variables aleatorias a partir de casi cualquier tipo de distribución de probabilidad.

Matemáticamente, la función de densidad uniforme se define como sigue:

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{fuera del intervalo } (a, b) \end{cases} \quad (4-17)$$

En esta expresión, X es una variable aleatoria definida en el intervalo (a, b) . La gráfica de la distribución uniforme queda ilustrada en la figura 4-5.

La función de la distribución acumulativa $F(x)$, para una variable aleatoria X uniformemente distribuida, se puede representar por

$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a} \quad 0 \leq F(x) \leq 1. \quad (4-18)$$

El valor esperado y la variancia de una variable aleatoria uniformemente distribuida están dados por las siguientes expresiones *

$$EX = \int_a^b \frac{1}{b-a} x dx = \frac{b+a}{2} \quad (4-19)$$

$$VX = \int_a^b \frac{(x-EX)^2}{b-a} dx = \frac{(b-a)^2}{12}. \quad (4-20)$$

Al efectuar aplicaciones de esta función, los parámetros de la función

* A fin de evitar confusiones con las variables FORTRAN que llevan subíndice, utilizamos los símbolos EX y VX para indicar, respectivamente, la media (o valor esperado) y a la variancia de X , en lugar de $E(X)$ y $V(X)$.

de densidad uniforme (4-17) esto es, los valores numéricos de a y de b , no necesariamente deben ser conocidos en forma directa. En casos típicos, aunque esto no sucede en todas las distribuciones uniformes, solamente conocemos la media y la variancia de la estadística que se va a generar. En estos casos, los valores de los parámetros se deben derivar al resolver el sistema que consta de las ecuaciones (4-19) y (4-20), para a y para b , pues se supone que EX y VX son conocidos. Este procedimiento, semejante

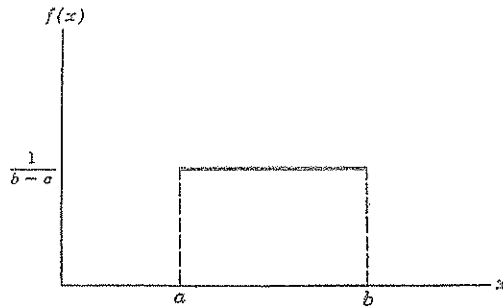


Figura 4-5

a una técnica de estimación conocida en la literatura estadística como método de momentos, proporciona las siguientes expresiones:

$$a = EX - \sqrt{3VX} \quad (4-21)$$

$$b = 2EX - a. \quad (4-22)$$

Para simular una distribución uniforme sobre cierto intervalo conocido (a, b) deberemos, en primer lugar, obtener la transformación inversa para la ecuación (4-18), de acuerdo con la ecuación (4-2).

$$x = a + (b - a)r \quad 0 \leq r \leq 1. \quad (4-23)$$

En seguida generamos un conjunto de números aleatorios correspondiente al rango de las probabilidades acumulativas, es decir, los valores de variables aleatorias uniformes definidas sobre el rango 0 a 1. Cada número aleatorio r determina, de manera única, un valor de la variable aleatoria x uniformemente distribuida.

Para aclarar estas afirmaciones es probable que lo mejor sea presentar una explicación gráfica. La figura 4-6 nos ilustra cómo cada valor generado de r está asociado con uno y sólo un valor de x . Por ejemplo, el valor específico de la función de distribución acumulativa en r_0 determina el valor de x en x_0 . Obviamente, este procedimiento se puede repetir tantas veces como se desee y en cada repetición se generará un nuevo valor de x . En este capítulo se observará cómo la generación de valores de variables aleatorias se puede seguir, mediante el uso de probabilidades acumulativas de muchas otras distribuciones. Aún más, esta técnica sirve de base para

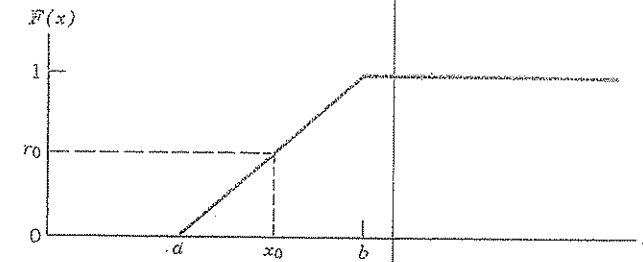


Figura 4-6

desarrollar los métodos más generales de los procesos de Monte Carlo, como se discutirá durante el desarrollo de este capítulo.

La figura 4-7 contiene un diagrama de flujo, asociado a la lógica que debe emplearse al simular una distribución uniforme para un intervalo dado (a, b) , para cuando se pretenda programar su estructura en una computadora. El diagrama de flujo se ha formulado de tal modo que resulta compatible con la subrutina FORTRAN, de la figura 4-8.

La primera declaración de nuestro diagrama FORTRAN para generar los valores de variables aleatorias, es una declaración CALL, que comprende el nombre de la subrutina y las variables de ésta. Los nombres de la subrutina usualmente se limitan a seis caracteres; por consiguiente, a esta subrutina en particular se le ha dado el nombre de UNIFORM. Los símbolos A , B y X que aparecen entre paréntesis representan a los parámetros a y b , así como al valor de la variable aleatoria uniforme x , respectivamente. El valor x se genera por medio de la subrutina y se regresa al programa principal mediante el efecto de la misma. Debe hacerse notar que la proposición CALL constituye una parte del programa principal, razón por lo que no se ha incluido en las instrucciones FORTRAN que aparecen en la figura 4-8.

La primera proposición en esta subrutina (figura 4-8) es de inicialización, la cual identifica, a su vez, la subrutina particular que es llamada por el programa principal. Cada una de las subrutinas que se describen en este capítulo

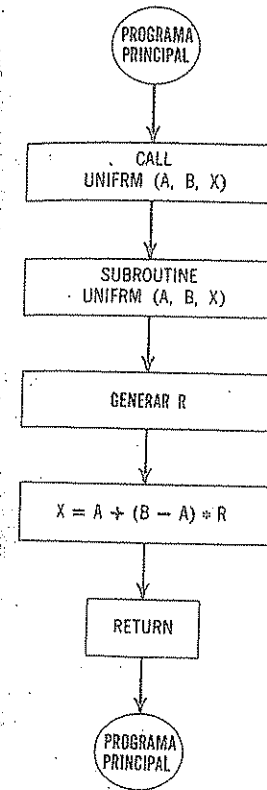


Figura 4-7. Diagrama de flujo de la generación de valores de variables aleatorias uniformes.

empezarán con una proposición similar de inicialización o de identificación.

La segunda proposición en la subrutina (figura 4-8) es una función de biblioteca cuyo efecto es el de fijar a la variable R igual a un número pseudoaleatorio generado por la función RND ; cada vez que se llama a la

1. SUBROUTINE UNIFORM (A, B, X)
2. $R = RND(R)$
3. $X = A + (B - A) * R$
4. RETURN

Figura 4-8. Subrutina FORTRAN para la generación de valores de variables aleatorias uniformes.

subrutina se genera un nuevo valor para R . La variable R es el símbolo FORTRAN que representa a r en la ecuación (4-23).

La tercera proposición en nuestra subrutina FORTRAN transforma a R del intervalo $(0, 1)$ al intervalo (a, b) mediante una expresión aritmética FORTRAN basada en la ecuación (4-23).

La cuarta proposición FORTRAN de la figura 4-8 regresa al programa principal el valor generado de X y el propio control del programa.

En lo que resta del presente capítulo se desarrollarán otras rutinas similares a la anterior.

La distribución exponencial

Durante nuestra experiencia diaria, observamos cómo transcurren los intervalos de tiempo definidos entre las ocurrencias de los eventos aleatorios distintos, y sobre la base de un plan de tiempos completamente independientes, recibimos información sobre numerosos eventos que ocurren en nuestro alrededor. Bastaría con citar los nacimientos, defunciones, accidentes, o conflictos mundiales, para mencionar sólo algunos. Si es muy pequeña la probabilidad de que ocurra un evento en un intervalo corto, y si la ocurrencia de tal evento es, estadísticamente independiente respecto a la ocurrencia de otros eventos, entonces el intervalo de tiempo entre ocurrencias de eventos de este tipo estará distribuido en forma exponencial. El hecho de que en el mundo real un proceso estocástico proporcione o no efectivamente valores de variables aleatorias de tipo exponencial, constituye una cuestión empírica cuya confirmación depende del grado en que satisfagan las suposiciones que sirven de base a la distribución exponencial. Específicamente, para los valores de variables aleatorias de tipo exponencial se deben satisfacer las siguientes suposiciones:

1. La probabilidad de que ocurra un evento en el intervalo $[t, (t + \Delta t)]$ es $\alpha \Delta t$.

2. α es una constante que no depende de t o de algún otro factor.

3. La probabilidad de que durante un intervalo $[t, (t + \Delta t)]$ ocurra más de un evento, tiende a 0 a medida que $\Delta t \rightarrow 0$, y su orden de magnitud deberá ser menor que el de $\alpha \Delta t$.

Curiosamente, se ha encontrado que el comportamiento de un considerable número de procesos dependientes del tiempo satisfacen las anteriores suposiciones un tanto fuertes. Por ejemplo, el intervalo entre los accidentes en una fábrica, la llegada de pedidos a una compañía, el registro de pa-

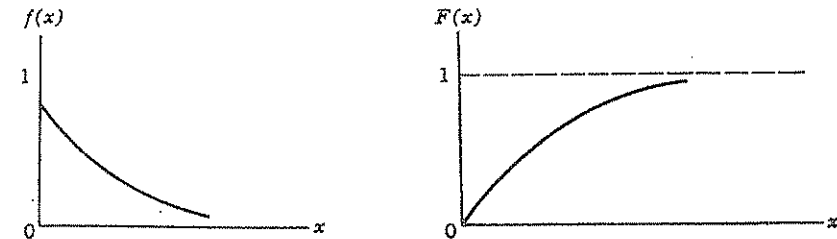


Figura 4-9

cientes en un hospital, el aterrizaje de aviones en un aeropuerto, etc., satisfacen una distribución exponencial.

Se dice que una variable aleatoria X tiene una distribución exponencial, si se puede definir a su función de densidad como:

$$f(x) = \alpha e^{-\alpha x} \quad (4-24)$$

con $\alpha > 0$ y $x \geq 0$.

La función de distribución acumulativa de X está dada por:

$$F(x) = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}, \quad (4-25)$$

y la media junto con la variancia de X se pueden expresar como

$$EX = \int_0^{\infty} x \alpha e^{-\alpha x} dx = \frac{1}{\alpha} \quad (4-26)$$

$$VX = \int_0^{\infty} \left(x - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha x} dx = \frac{1}{\alpha^2} = (EX)^2 \quad (4-27)$$

En la figura 4-9 aparece la gráfica de la distribución exponencial.

Obsérvese que, como la distribución exponencial solamente tiene un parámetro α , es posible expresarlo como:

$$\alpha = \frac{1}{EX} \quad (4-28)$$

Existen muchas maneras para lograr la generación de valores de variables

aleatorias exponenciales. Puesto que $F(x)$ existe explícitamente, la técnica de la transformación inversa nos permite desarrollar métodos directos para dicha generación. Debido a la simetría que existe entre la distribución uniforme sigue que la intercambiabilidad de $F(x)$ y $1-F(x)$. Por lo tanto

$$r = e^{-\alpha x} \quad (4-29)$$

y consecuentemente

$$x = -\left(\frac{1}{\alpha}\right) \log r = -EX \log r. \quad (4-30)$$

Por consiguiente, para cada valor del número pseudoaleatorio r se determina un único valor para x . Los valores de x toman tan sólo magnitudes no negativas, debido a que $\log r \leq 0$ para $0 \leq r \leq 1$, y además se ajustan a la función de densidad exponencial (4-24) con un valor esperado EX . Conviene hacerle notar al lector, que pese a que esta técnica parece en principio muy simple, es preciso recordar que en una computadora digital el cálculo del logaritmo natural involucra una expansión en serie de potencias (o una técnica equivalente de aproximación) para cada valor de variable aleatoria uniforme que se deba generar.

La figura (4-10) contiene un diagrama de flujo para generar valores de variables aleatorias exponenciales, mientras que la figura 4-11 contiene la subrutina FORTRAN correspondiente. El nombre que se da a esta subrutina es EXPENT.

La técnica de la transformación inversa no es el único método para generar valores de variables aleatorias exponenciales con procedimientos internos. En este libro se esbozarán dos métodos adicionales: el motivado por su importancia histórica, y el debido a su rapidez potencial.

Un ejemplo ingenioso de utilización de la técnica de rechazo lo constituye el método de Von Neumann [29] para generar valores de variables aleatorias exponenciales; resulta una de esas penosas realidades de la vida, que el empleo de la transformación logarítmica, para el mismo propósito, sea más conveniente, de

1. SUBROUTINE EXPENT (EX, X)
2. R = RND (R)
3. X = -EX * LOG (R)
4. RETURN

Figura 4-11. Subrutina FORTRAN para la generación de valores de variable aleatoria exponencial.

vido a una ligera diferencia en rapidez. Debido a que este método se menciona muy a menudo en la literatura, aquí ofrecemos una breve descripción de su estructura.

Al generar los números aleatorios uniformes $r_1, r_{11}, r_{12}, \dots, r_{1j}, r_2, r_{21}, r_{22}, \dots$ y tomar la sucesión de sumas:

$$1 - r_1 + \sum_{i=1}^j r_{1i},$$

$$1 - r_2 + \sum_{i=1}^j r_{2i},$$

$$\dots \dots \dots$$

$$1 - r_t + \sum_{i=1}^j r_{ti}.$$

cada una de estas series, se acorta o termina, para aquel valor de j para el cual $1 - r_t + \sum_{i=1}^j r_{ti} \geq 1$ por primera vez. La sucesión se termina tan pronto como el primer valor de t implique un valor impar para j . La cantidad

$$x = (t - 1) + r_t \quad (4-31)$$

constituye un valor de variable aleatoria exponencial para el cual $EX = 1$. En [7, p. 262] y [5, p. 166], se presentan, respectivamente, una prueba del método y un diagrama de flujo para su programación.

El segundo método que se delinea aquí, debido a G. Marsaglia [18], también sirve para generar valores de variable aleatoria con distribución exponencial con media y variancia unitarias, sin requerir los beneficios de la transformación logarítmica. Si el programa correspondiente se escribe en un lenguaje similar al de la subrutina para calcular logaritmos, este método resulta más rápido que el de la técnica de transformación inversa. Además, este procedimiento constituye un ejemplo excelente del método de composición.

El valor de la variable aleatoria x distribuida exponencialmente, está dado por

$$x = m + \min_i (r_1, r_2, \dots, r_i, \dots, r_n) \quad (4-32)$$

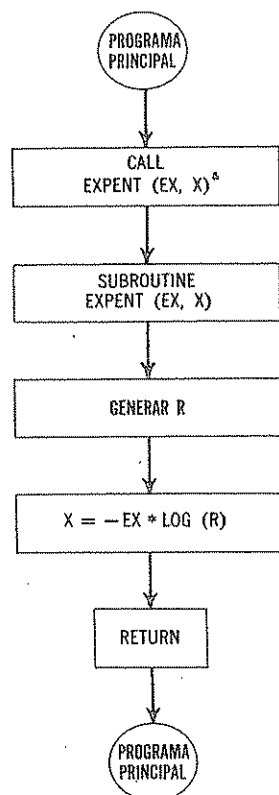


Figura 4-10. Diagrama de flujo para la generación de valores de variable aleatoria con distribución exponencial.

donde las r_i son números pseudoaleatorios uniformemente distribuidos, los valores de m y n para una x particular están determinados por las siguientes distribuciones acumulativas discretas de probabilidad:

$$P(M \leq m) = \sum_{k=0}^m \frac{1}{ce^{k+1}} \quad \text{para } m = 0, 1, 2, \dots \quad (4-33)$$

$$P(N \leq n) = \sum_{k=1}^n \frac{c}{k!} \quad \text{para } n = 1, 2, 3, \dots \quad (4-34)$$

donde $c = 1/(e - 1)$. Los valores (de variables aleatorias) de m y n computan mediante la generación de dos números aleatorios, cuyos valores correspondientes se calculan haciendo uso de la transformación inversa en las ecuaciones (4-33) y (4-34). A fin de obtener un valor de variable aleatoria exponencial la probabilidad de que sólo se generen tres números aleatorios es de 0.58; para no más de cinco, la probabilidad es de 0.99. Cuando se desee generar valores de variables aleatorias con $EX \neq 1$, bastará con multiplicar el valor de x , determinado por la ecuación (4-33) por el valor de EX .

La distribución exponencial se basa en el supuesto de que se tiene un parámetro constante α . En otras palabras, se supone que todos los eventos han sido generados por un solo proceso aleatorio. En el mundo real frecuentemente se viola esta suposición, sobre todo, cuando tratamos con eventos indistinguibles que son producto de procesos aleatorios diferentes pero entremezclados. Resulta en muchas ocasiones posible, y hasta frecuente que al tomar una muestra ésta provenga de dos o más distribuciones exponenciales, cada una de las cuales tenga un valor diferente de α . En este caso tenemos una mayoría de problemas sobre fenómenos de espera (colas); por ejemplo, las llegadas pueden ocurrir en razones iguales a α_i con probabilidades p_i , en las que α_i representa al parámetro de la i -ésima población ($i = 1, 2, \dots, s$), tal que $\alpha_i \neq \alpha_j$ y $\sum_{i=1}^s p_i = 1$. Esta mezcla de valores exponenciales se dice que obedece a una distribución hipereexponencial ($s = 2$) o a una exponencial generalizada ($s > 2$). Para generar esta mezcla dada de s valores de variables aleatorias exponenciales, sencillamente se intercala un interruptor de probabilidad antes de la instrucción CALL en la figura 4-10, con lo cual se puede decidir cuál de los s valores de $1/\alpha_i$, previamente almacenados, debe ser el que se aplique en la transformación inversa. El problema real radica en el hecho de que raramente conocemos cuál es el tipo de mezcla que se tiene o debe generar. Sin embargo, algunas veces se pueden considerar ciertas hipótesis simplificadas. Dos casos de este tipo se describen a continuación:

1. Las llegadas se originan solamente en dos poblaciones y tienen probabilidades p y $(1 - p)$ con parámetros $p2\alpha$ y $(1 - p)2\alpha$, respectivamente.

Este proceso genera valores de variable aleatoria hipereexponenciales [20] con media igual a $1/\alpha$ y con una función de densidad:

$$f(x) = 2p^2\alpha \exp(-2p\alpha x) + 2(1 - p)^2\alpha \exp[2(1 - p)\alpha x] \quad (4-35)$$

La variancia de x es

$$VX = \frac{1}{\alpha^2} \left[\frac{1}{2p(1 - p)} - 1 \right], \quad (4-36)$$

la cual indica que los valores hipereexponenciales siempre tienen una variancia mayor de $1/\alpha^2$, a menos que $p = 1/2$. En caso de conocer el valor de la sobredispersión $VX/(EX)^2$, para cierto valor dado de $EX = 1/\alpha$ entonces se puede determinar a p como sigue:

$$p = \frac{1}{2} - \frac{1}{2} \left[1 - \frac{2}{VX/(EX)^2 + 1} \right]^{1/2} \quad (4-37)$$

y usarse en el programa como se indicó previamente.

2. Las llegadas se originan en un número infinito de poblaciones diferentes, con un número α también diferente para cada una. Debido a que α es siempre positiva y no está restringida a valores enteros, podemos suponer que la probabilidad de cualquier valor de α se ajusta a una distribución gamma. En los casos en que componemos valores de variables aleatorias, distribuidos exponencialmente y que se ajustan a una distribución gamma, diremos que se forma una distribución *exponencial generalizada* (o bien una distribución Pearson del tipo XI) [16]. La función de densidad correspondiente es:

$$f(x) = k\alpha^k (a + x)^{-(k+1)}, \quad (4-38)$$

y su función de distribución acumulativa está dada por:

$$F(x) = 1 - \left(\frac{a}{a + x} \right)^k, \quad (4-39)$$

donde k y a son los parámetros de la variable aleatoria repartida de acuerdo con la distribución gamma. Haciendo $r = 1 - F(x)$ y aplicando la transformación inversa de la expresión dada en (4-39), se sigue que:

$$x = a \left[\left(\frac{1}{r} \right)^{1/k} - 1 \right] \quad (4-40)$$

es una variable aleatoria con una función de densidad igual a la que se expresa en la ecuación (4-38). Para generar valores de variable aleatoria que se ajusten a una distribución exponencial con media EX y variancia

VX , tales que $VX > (EX)^2$, se pueden utilizar las siguientes fórmulas:

$$k = \frac{2VX}{[VX - (EX)^2]} \quad (4-41)$$

$$a = (k - 1)EX, \quad (4-42)$$

las cuales resultan válidas solamente para valores de $k > 2$.

La generalización de la distribución exponencial resulta una tarea demasiado simple cuando se consideran variables positivas $x \geq a$ con a estrictamente positiva. En este caso, la substitución de $(x - a)$ por x conduce de inmediato a la así llamada distribución exponencial no central o de doble parámetro, la cual se puede simular sin dificultad alguna por cualquiera de los métodos que se han tratado previamente.

Una más de las extensiones de la distribución exponencial es la conocida por el nombre de distribución Weibull [30], con una función de densidad y una distribución acumulativa que se pueden expresar como sigue:

$$f(x) = \frac{c}{b} \left(\frac{x-a}{b} \right)^{c-1} \exp \left[- \left(\frac{x-a}{b} \right)^c \right] \quad (4-43)$$

$$F(x) = 1 - \exp \left[- \left(\frac{x-a}{b} \right)^c \right] \quad (4-44)$$

para $x \geq a$, $a \geq 0$, $b > 0$ y $c > 0$.

En esta descripción el parámetro de ubicación a se supone con un valor idénticamente nulo, lo cual deja a b y c , como parámetros de escala y de forma, respectivamente. Como únicos datos de caracterización, el papel que desempeña la constante $1/b$ es análogo al de α en la distribución exponencial. Si $c = 1$, la expresión (4-43) resulta idéntica a la dada en la ecuación (4-24), pero si $c > 1$, entonces la distribución adopta la forma de una campana. En cualquier otro caso esta distribución mantendrá la forma de una "J", al igual que la distribución exponencial. Se ha encontrado que esta distribución resulta de gran utilidad cuando se realiza el tratamiento de problemas relacionados con vida media, resistencia a la ruptura y los datos relacionados con los fenómenos de confiabilidad, para los que su comportamiento ha sido descrito satisfactoriamente.

El valor esperado de la ecuación (4-43) está dado por

$$EX = b\Gamma \left(\frac{1}{c} + 1 \right), \quad (4-45)$$

y su variancia

$$VX = b^2\Gamma \left(\frac{2}{c} + 1 \right) - (EX)^2. \quad (4-46)$$

La transformación inversa de la ecuación (4-44), utilizando $r = 1 - F(x)$, conduce al proceso de generación de valores de variable aleatoria con distribución Weibull, mediante la siguiente expresión:

$$x = b(-\log r)^{1/c}. \quad (4-47)$$

La distribución gamma

Si un determinado proceso consiste de k eventos sucesivos y si el total del tiempo transcurrido para dicho proceso se puede considerar igual a la suma de k valores independientes de la variable aleatoria con distribución exponencial, cada uno de los cuales tiene un parámetro definido α , la distribución de esta suma coincidirá con una distribución gamma con parámetros α y k . La suma de los k (donde k es un entero positivo) valores de variable aleatoria con distribución exponencial con un mismo parámetro α , también recibe el nombre de distribución de Erlang [4]. Desde un punto de vista matemático la distribución de Erlang no es otra cosa que la convolución de k distribuciones exponenciales, o sea la distribución de la suma de k variables exponenciales. Aun más, si k se adecúa a la distribución binomial negativa o a la distribución geométrica [6], entonces la suma de k valores de variable aleatoria, con la misma α , adoptará la forma de una distribución gamma. La forma más general de una distribución gamma se obtiene al considerar a k con valores positivos pero sin que estén restringidos a ser enteros. Casi siempre resulta posible ajustar alguna de las formas de la distribución gamma a un buen número de distribuciones de datos estadísticos sesgados en forma positiva.

La función gamma está descrita mediante la siguiente función de densidad:

$$f(x) = \frac{\alpha^k x^{(k-1)} e^{-\alpha x}}{(k-1)!} \quad (4-48)$$

donde $\alpha > 0$, $k > 0$ y x se considera no negativo. Pese a que no existe una forma explícita para describir la función acumulativa de la distribución gamma, Pearson [31] ha logrado presentar, en forma tabular, los valores de la llamada función gamma incompleta. Respecto a la media y la variancia de esta distribución, sus correspondientes expresiones están formuladas como sigue:

$$EX = \frac{k}{\alpha} \quad (4-49)$$

$$VX = \frac{k}{\alpha^2}. \quad (4-50)$$

Si $k = 1$, entonces la distribución gamma resulta ser idéntica a la distribución exponencial; mientras que si k es un entero positivo, la distri-

bución gamma coincide con la distribución de Erlang. Conviene también anotar que, a medida que k se incrementa, la distribución gamma tiende, en forma asintótica, a una distribución normal.

Para generar valores de variable aleatoria con distribución gamma y con un valor esperado y variancia dados, se pueden emplear las siguientes fórmulas a fin de determinar los parámetros de $f(x)$ en la ecuación (4-48):

$$\alpha = \frac{EX}{\sqrt{X}} \quad (4-51)$$

$$k = \frac{(EX)^2}{\sqrt{X}} \quad (4-52)$$

Debido a que para una distribución gamma no se puede formular explícitamente una función de distribución acumulativa, debemos considerar un método alternativo para generar valores de variable aleatoria con distribución gamma. En relación a tales valores que satisfacen la distribución Erlang, éstos se pueden generar con sólo reproducir el proceso aleatorio sobre el cual se basa la distribución de Erlang. Para lograr este resultado se debe tomar la suma de los k valores de variable aleatoria con distribución exponencial x_1, x_2, \dots, x_k , cuyo valor esperado es el mismo e igual a $1/\alpha$. En consecuencia, el valor de la variable aleatoria (Erlang) x se puede expresar como

$$x = \sum_{i=1}^k x_i = -\frac{1}{\alpha} \sum_{i=1}^k \log r_i \quad (4-53)$$

En las figuras 4-12 y 4-13 aparece un diagrama de flujo y una subrutina en FORTRAN que sirve para generar una distribución Erlang (gamma), en la que k toma valores enteros. El lector debe notar que la ecuación (4-53) no se utiliza en la codificación del programa FORTRAN; en lugar de ésta, se utiliza una forma computacional equivalente y más rápida, que la substituye. Esta expresión corresponde a la siguiente fórmula:

$$x = -\frac{1}{\alpha} \left(\log \prod_{i=1}^k r_i \right) \quad (4-54)$$

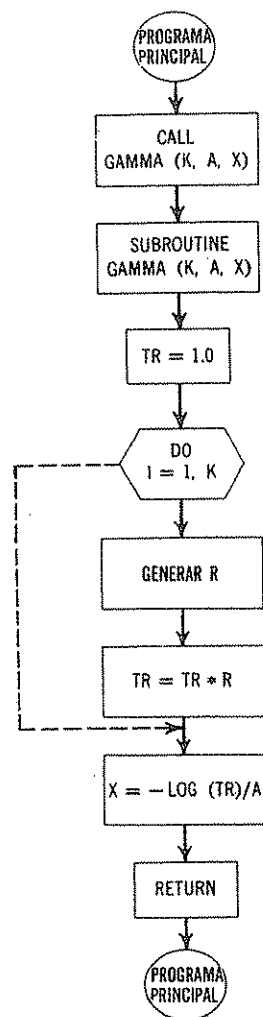


Figura 4-12. Diagrama de flujo para generar valores de variable aleatoria con distribución gamma.

obsérvese también que en la subrutina se utilizan A y K para denotar a α y k , respectivamente.

El problema de generar valores de variable aleatoria con distribución gamma cuando el parámetro k no es un entero, queda aún en la actividad como un problema por resolverse, ya que en este caso todavía se formula un método estocástico satisfactorio, sin lo cual no puede justificarse algún proceso de simulación correspondiente. Empero, delinearemos uno de los intentos producidos para resolver esta dificultad. Si k es un número racional, entonces se lo puede expresar mediante la suma de un entero más una fracción, de modo tal que $k = k_1 + q$, con $0 < q < 1$; más aún, si $k_2 = k_1 + 1$, entonces se tiene que $k_2 - k_1 = 1 - q$. Nótese que aquí sólo consideraremos el caso en que $k > 1$. Puesto que el valor esperado, la va-

1. SUBROUTINE GAMMA (K, A, X)
2. TR = 1.0
3. DO 5 I = 1, K
4. R = RND (R)
5. TR = TR * R
6. X = -LOG (TR)/A
7. RETURN

Figura 4-13. Subrutina FORTRAN para la generación de valores de variable aleatoria con distribución gamma.

riancia y el tercer momento central de k [23], la distribución gamma de parámetro k se puede aproximar con sólo considerar una mezcla adecuada de valores de variable aleatoria con distribución gamma, para la cual se eligen valores de k_2 con probabilidad q y valores de k_1 con probabilidad $1 - q$. Si los valores de k son suficientemente grandes, entonces la aproximación lograda con este criterio proporcionará resultados más satisfactorios. Si utilizamos esta técnica en el programa FORTRAN mencionado debemos insertar un interruptor de probabilidad que defina unívocamente los valores de k , antes de la proposición DO (en la figura 4-13).

Relacionadas con las variables gamma existe un buen número de distribuciones de probabilidad, de las cuales las más importantes que se pueden mencionar son: la distribución Ji cuadrada y la Beta.

La distribución Ji cuadrada no es otra cosa que una distribución gamma, para la cual $\alpha = 1/2$. Consecuentemente, las variables con distribución Ji cuadrada tienen una media igual a $EX = 2k$, que recibe el nombre de grados de libertad, y una variancia $VX = 4k$. En el caso de ser EX un número par, entonces k es un entero y, en consecuencia, se podrá aplicar la técnica de generación correspondiente a la ecuación (4-54). Si por el

contrario, EX resulta un número impar, entonces $k = EX/2 - 1/2$, y por lo tanto:

$$x = -\frac{1}{\alpha} \log \left(\prod_{i=1}^k r_i \right) + z^2, \quad (4-55)$$

donde z^2 es el cuadrado de un valor de variable aleatoria con distribución normal, la cual tiene una media nula y una variancia unitaria. La generación de valores de variable aleatoria normales se describirá en la sección siguiente.

La distribución Beta corresponde a los cocientes que se producen con dos variables gamma x_1 y $(x_1 + x_2)$, donde x_1 y x_2 son dos variables independientes gamma a las que corresponde un mismo valor de los parámetros α , k_1 y k_2 , respectivamente, de modo que $k = (k_1 + k_2)$ resulte el parámetro de $(x_1 + x_2)$. La variable Beta está dada por

$$x = \frac{x_1}{x_1 + x_2} \quad 0 < x < 1. \quad (4-56)$$

Por lo tanto, para generar un valor de variable aleatoria x con distribución Beta, se debe obtener el cociente de dos valores de variable aleatoria con distribución gamma, uno con parámetro k y el otro con parámetro k .

La función de densidad de la distribución Beta es

$$f(x) = \frac{\Gamma(a+b)x^{a-1}(1-x)^{b-1}}{\Gamma(a)\Gamma(b)} \quad (4-57)$$

con un valor esperado y una variancia iguales a

$$EX = a/(a+b) \quad (4-58)$$

$$VX = \frac{(EX)(b)}{(a+b+1)(a+b)}, \quad (4-59)$$

donde a y b corresponden a los parámetros k_1 y k_2 de los valores de variables aleatorias con distribución gamma, que cumplen la ecuación (4-56). Esta distribución Beta tiene una amplia aplicación, principalmente en los programas referentes al empleo de las técnicas de camino crítico, como por ejemplo PERT [9]. Con estos mismos principios se puede obtener una versión generalizada de la distribución Beta, llamada distribución de Dirichlet, para la cual basta considerar en el modelo más de dos cocientes no constantes. Este caso suele surgir al simular vectores fijos o renglones de matrices estocásticas cuyas entradas están asociadas a situaciones de probabilidad variable, así como también, cuando se modelan composiciones aleatorias de ciertas unidades físicas que se encuentran subdivididas en partes o componentes distribuidas al azar.

La distribución normal

La más conocida y más ampliamente utilizada distribución de probabilidad es sin duda la distribución normal y su popularidad se debe cuando menos a dos razones que presentan sus propiedades generales. "Las pruebas matemáticas nos señalan, que *bajo ciertas condiciones de calidad*, resulta justificado que esperemos una distribución normal mientras que la experiencia estadística muestre que, de hecho, muy a menudo las distribuciones se aproximan a la normal" [8, p. 232].

La distribución normal basa su utilidad en el teorema del límite central. Este teorema postula que, la distribución de probabilidad de la suma de N valores de variable aleatoria x_i independientes pero idénticamente distribuidos, con medias respectivas μ_i y variancias σ_i^2 se aproxima asintóticamente a una distribución normal, a medida que N se hace muy grande, y que dicha distribución tiene como media y variancia respectivamente, a

$$\mu = \sum_{i=1}^N \mu_i \quad (4-60)$$

$$\sigma^2 = \sum_{i=1}^N \sigma_i^2. \quad (4-61)$$

En consecuencia, el teorema del límite central permite el empleo de distribuciones normales para representar medidas globales operadas sobre los efectos de causas (errores) aditivas distribuidas en forma independiente, sin importar la distribución de probabilidad a que obedezcan las mediciones de causas individuales [25, p. 262]. El carácter que se atribuye a esta forma de interpretación del teorema del límite central resulta de particular importancia, ya que de esta manera se provee una justificación matemática para manipular la evidencia empírica, que tan frecuentemente aparece en los datos distribuidos en forma aproximadamente normal, obtenidos en una gran mayoría de problemas de investigación.

Otro rasgo valiosamente práctico de la distribución normal, es su utilidad para aproximar distribuciones de Poisson y binomiales para mencionar sólo dos entre muchas. A partir de la distribución normal, se pueden derivar otras muchas de las distribuciones existentes que juegan un papel muy importante en la estadística moderna, por ejemplo la Ji cuadrada, la t , y la distribución F , las cuales se originan a partir de consideraciones hechas sobre la distribución de probabilidad de la suma de los cuadrados de un número específico de valores de variables aleatoria con una distribución normal estándar [8, p. 233].

Si la variable aleatoria X tiene una función de densidad $f(x)$ dada como

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-1/2 \left(\frac{x - \mu_x}{\sigma_x} \right)^2}, \quad -\infty < x < \infty, \quad (4-62)$$

con σ_x positiva, entonces se dice que X tiene una distribución normal o Gaussiana, con parámetros μ_x y σ_x . En la figura 4-14 se muestra la ya conocida gráfica con su forma de campana, de la función de densidad normal.

Si los parámetros de la distribución normal tienen los valores $\mu_x = 0$ y $\sigma_x = 1$, la función de distribución recibirá el nombre de *distribución normal estándar*, con función de densidad

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad -\infty < z < \infty. \quad (4-63)$$

Cualquiera distribución normal se puede convertir a la forma estándar, mediante la siguiente substitución:

$$z = \frac{x - \mu_x}{\sigma_x}. \quad (4-64)$$

La función de distribución acumulativa $F(x)$ o $F(z)$ no existe en forma explícita; sin embargo, esta última se encuentra totalmente tabulada

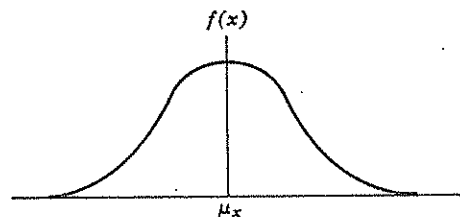


Figura 4-14. Una distribución normal.

en cualquier libro sobre estadística. El valor esperado y la variancia de la distribución normal no estándar están dados por:

$$EX = \mu_x \quad (4-65)$$

$$VX = \sigma_x^2. \quad (4-66)$$

Existen varios métodos para generar en una computadora, valores de variable aleatoria distribuidos en forma normal. Debido a su popularidad sólo discutiremos y demostraremos en detalle el procedimiento llamado del límite central. También se incluirán breves descripciones de otros procedimientos.

A fin de simular una distribución normal con media μ_x y variancia σ_x^2 dadas, se debe proponer la siguiente interpretación matemática del teorema del límite central. Si r_1, r_2, \dots, r_N representan variables aleatorias indepen-

dientes, cada una de las cuales posee la misma distribución de probabilidad caracterizada por $E(r_i) = \theta$ y $\text{Var}(r_i) = \sigma^2$, entonces

$$\lim_{N \rightarrow \infty} P \left[a < \frac{\sum_{i=1}^N r_i - N\theta}{\sqrt{N}\sigma} < b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz, \quad (4-67)$$

donde

$$E \left(\sum_{i=1}^N r_i \right) = N\theta, \quad (4-68)$$

$$\text{Var} \left(\sum_{i=1}^N r_i \right) = N\sigma^2, \quad (4-69)$$

$$z = \frac{\sum_{i=1}^N r_i - N\theta}{\sigma \sqrt{N}}. \quad (4-70)$$

Tanto de la definición de la distribución normal estándar como de la ecuación (4-64), se sigue que z es un valor de variable aleatoria con distribución normal estándar.

El procedimiento para simular valores normales utilizando computadoras requiere el uso de la suma de K valores de variable aleatoria distribuidos uniformemente; esto es, la suma de r_1, r_2, \dots, r_K , con cada r_i definida en el intervalo $0 < r_i < 1$. Aplicando la convención notacional de la forma matemática del teorema del límite central, así como nuestros conocimientos previos de la distribución uniforme, encontramos que

$$\theta = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}, \quad (4-71)$$

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{1}{\sqrt{12}}, \quad (4-72)$$

$$z = \frac{\sum_{i=1}^K r_i - K/2}{\sqrt{K/12}}. \quad (4-73)$$

Pero, por definición, z es un valor de variable aleatoria con distribución normal estándar que se puede escribir en la forma sugerida por la ecuación (4-64), donde x es un valor de variable aleatoria distribuido en forma normal que se va a simular, con media μ_x y variancia σ_x^2 . Igualando las

ecuaciones (4-73) y (4-64) obtendremos:

$$\frac{x - \mu_x}{\sigma_x} = \frac{\sum_{i=1}^K r_i - K/2}{\sqrt{K/12}}, \quad (4-74)$$

y resolviendo para x , se tiene que

$$x = \sigma_x \left(\frac{12}{K} \right)^{1/2} \left(\sum_{i=1}^K r_i - \frac{K}{2} \right) + \mu_x. \quad (4-75)$$

Por lo tanto, mediante la ecuación (4-75) podemos proporcionar una formulación muy simple para generar valores de variable aleatoria normalmente distribuidos, cuya media sea igual a μ_x y variancia σ_x^2 . Para generar un solo valor de x (un valor de variable aleatoria con distribución normal) bastará con sumar K números aleatorios definidos en el intervalo de 0 a 1. Substituyendo el valor de esta suma en la ecuación (4-75), así como también los valores de μ_x y σ_x para la distribución deseada, encontraremos que se ha determinado un valor particular de x . Ciertamente, este procedimiento se puede repetir tantas veces como valores de variable aleatoria normalmente distribuidos se requieran.

El valor de K que debe aplicarse a las fórmulas usualmente se determina al establecer las condiciones de balance entre la eficiencia de cómputo y la precisión. Al considerar la convergencia asintótica implicada por el procedimiento del límite central, es deseable que K corresponda a un número muy grande. Considerando el tiempo que comprende la generación de K valores uniformes por cada valor de variable aleatoria normal, sería preferible que K estuviera asociada a un número muy chico. En la práctica de simulación se recomienda una $K = 10$, como el menor valor deseable. Sin embargo, con $K = 12$ se logra una cierta ventaja computacional, ya que en la ecuación (4-75) se puede evitar una multiplicación constante. No obstante, este valor de K trunca la distribución a los límites ± 6 , y además, se ha encontrado que no es confiable para valores de x mayores que tres de las desviaciones estándar, aunque pese a lo anterior, la experiencia muestra que este criterio conduce a programas de razonable rapidez [21, p. 381]. Con el fin de obtener mayor precisión se deben considerar valores mayores para K (del orden de $K = 24$), de acuerdo con la llamada técnica de aproximación de Teichroew [27], aunque en estos casos la eficiencia del procedimiento del límite central resulta significativamente menor a la de otros criterios.

La aproximación de Teichroew mejora la precisión de las probabilidades de los eventos extremos, obtenidas con el procedimiento del límite central. Con $K = 12$ deberemos calcular el valor de

$$y = \frac{\left(\sum_{i=1}^{12} r_i - 6 \right)}{4}, \quad (4-76)$$

y el siguiente polinomio servirá para obtener el valor de la variable aleatoria z distribuida normalmente:

$$z = a_1 y + a_3 y^3 + a_5 y^5 + a_7 y^7 + a_9 y^9, \quad (4-77)$$

donde

$$a_1 = 3.949846138$$

$$a_3 = 0.252408784$$

$$a_5 = 0.076542912$$

$$a_7 = 0.008355968$$

$$a_9 = 0.029899776$$

Pese a que el método de aproximación de Teichroew se puede programar fácilmente en FORTRAN, nuestro diagrama de flujo (figura 4-15) y subrutina FORTRAN (figura 4-16) para generar valores de variable aleatoria con distribución normal, con un valor dado tanto para el valor esperado como para la variancia, sigue el criterio del límite central. En la subrutina se utiliza la siguiente notación

$$EX = \mu_x \quad (4-78)$$

$$STD X = \sigma_x \quad (4-79)$$

$$SUM = \sum R. \quad (4-80)$$

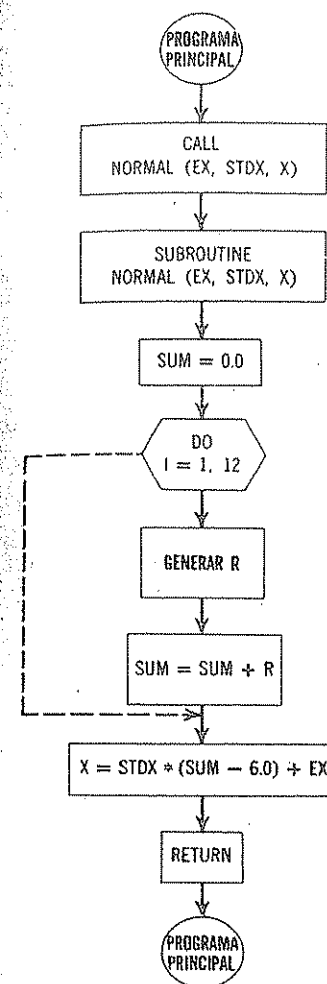


Figura 4-15. Diagrama de flujo para la generación de valores de variable aleatoria con distribución normal.

Ahora incluiremos un breve resumen de otros dos procedimientos para generar valores de variables aleatorias distribuidas normalmente.

```

1. SUBROUTINE NORMAL (EX, STD, X)
2. SUM = 0.0
3. DO 5 I = 1, 12
4. R = RND (R)
5. SUM = SUM + R
6. X = STD * (SUM - 6.0) + EX
7. RETURN

```

Figura 4-16. Subrutina FORTRAN para generar valores de variable aleatoria con distribución normal.

El procedimiento directo [3]

Sean r_1 y r_2 dos valores de variable aleatoria independientes distribuidos de modo uniforme y definidos en el intervalo $(0, 1)$, entonces:

$$x_1 = (-2 \log_e r_1)^{1/2} \cos 2\pi r_2 \quad (4-81)$$

$$x_2 = (-2 \log_e r_1)^{1/2} \sin 2\pi r_2, \quad (4-82)$$

serán dos valores de variable aleatoria obtenidos a partir de una distribución normal estándar. Con este método se logran resultados exactos y su velocidad de cómputo se puede comparar con la del método del límite central, sujeto a la eficiencia de ciertas subrutinas para funciones especiales [2], p. 382].

El procedimiento rápido [19]

Se afirma que esta técnica es la más rápida, aunque se le imputa que emplea varios cientos de ubicaciones de memoria para almacenar en la computadora ciertas constantes específicas. Las desviaciones aleatorias normales se calculan a partir de la mezcla de tres densidades:

$$f(x) = 0.9578g_1(x) + 0.0395g_2(x) + 0.0027g_3(x). \quad (4-83)$$

Los fundamentos racionales en los que se basa esta mezcla establecen que del 95 al 87 por ciento del tiempo sólo se usa $g_1(x)$, lo cual proporciona inmediatamente una variable normal con una tabla de longitud mínima. Las otras dos funciones son considerablemente más complicadas; si se desean mayores detalles o información sobre las constantes que emplea este procedimiento, se encontrarán en la referencia que se cita.

También pueden emplearse otros caminos, para generar valores de variable aleatoria normales, como el *método de rechazos* (Von Neumann) y el *método de Hastings*, aunque ninguno de estos métodos ofrece más simplicidad, precisión o rapidez, que los vistos previamente.

Las siguientes distribuciones se derivan de la distribución normal y por lo tanto, se pueden simular en forma indirecta por medio de ciertas funciones cuyos argumentos sean valores de variable aleatoria con distribución normal.

La distribución Ji cuadrada se obtiene mediante la suma de los cuadrados de los valores de variable aleatoria independientes, que obedecen a una distribución normal estándar. La variable Ji cuadrada se denota por

$$x = \chi_m^2 = \sum_{i=1}^m z_i^2, \quad (4-84)$$

donde z_i es una variable normal estándar y m representa los *grados de libertad*. La función de densidad χ_m^2 o x es una densidad gamma (véase la ecuación 4-48), con $k = m/2$ y $\alpha = 1/2$. En consecuencia,

$$f(x) = \frac{x^{\frac{m}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{m}{2}} \left(\frac{m}{2} - 1\right)!}, \quad (4-85)$$

y $EX = m$ y $VX = 2m$.

Esta distribución queda completamente descrita una vez conocidos los grados de libertad y se pueden, por lo tanto, simular muy fácilmente mediante la ecuación (4-54) si m resulta par. Cuando se trate de grados de libertad impares y $m < 30$, pueden utilizarse las fórmulas (4-55) o (4-84); mientras que si $m > 30$ se podrá emplear el criterio de la aproximación normal para variables Ji cuadrada, basado en la conocida fórmula:

$$z = \sqrt{2\chi_m^2} - \sqrt{2m} - 1. \quad (4-86)$$

Resolviendo para el valor Ji cuadrado χ_m^2 , obtenemos:

$$x = \chi_m^2 = \frac{(z + \sqrt{2m - 1})^2}{2}. \quad (4-87)$$

La distribución t describe una variable aleatoria tal, que

$$t = \frac{z}{\sqrt{\chi_m^2/m}} \quad (4-88)$$

con una función de densidad

$$f(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \left[\pi m \left(1 + \frac{t^2}{m}\right)\right]^{1/2}} \quad (4-89)$$

La distribución acumulativa correspondiente a la distribución t se encuentra disponible tan sólo en forma tabular; $EX = 0$ y $VX = m/(m-2)$.

Para propósitos de simulación se puede usar la fórmula (4-88) o sea, el cociente que se obtiene entre los valores de variable aleatoria con distribución normal estándar y los valores de variable aleatoria distribuidos en forma Ji cuadrada; en forma alternativa, es posible utilizar un valor de variable aleatoria distribuida en forma normal estándar, cuya variancia sea $m/(m-2)$. Para $m > 30$, bastará con que se aplique el criterio directo de la aproximación normal.

La distribución F corresponde a la distribución de probabilidad de los cocientes que se obtienen entre las sumas correspondientes de los cuadrados de los valores de variable aleatoria con distribución normal. En otras palabras, se puede representar la variable F bajo la siguiente notación

$$F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n}, \quad (4-90)$$

donde m y n representan los correspondientes grados de libertad de las sumas independientes, al igual que los valores Ji cuadrados. Tanto la función de densidad como la acumulativa de las variables F son lo suficientemente complicadas como para incluirlas en este libro; sin embargo, sus momentos son relativamente simples [31, p. 187], es decir,

$$EX = \frac{n}{n-2}, \quad n > 2 \quad (4-91)$$

$$VX = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n > 4. \quad (4-92)$$

Finalmente, conviene anotar que los valores F se pueden simular mediante el uso de las ecuaciones (4-90) y (4-84) o bien, con la ecuación (4-54).

La distribución normal multivariada

Esta distribución está definida por un vector de variables aleatorias para el que cada una de sus componentes representa una variable aleatoria normal, con una media y variancia conocidas. Cuando el conjunto de componentes de este vector normal aleatorio son independientes entre sí la generación de vectores aleatorios normales se seguirá en forma directa de las técnicas ya discutidas en la sección anterior. Pero si las componentes del vector aleatorio normal son dependientes, entonces, las covariancias entre las variables que representan las componentes serán distintas de cero, y será indispensable el uso de la matriz variancia-covariancia para lograr la generación de vectores aleatorios normales.

Denotaremos por \mathbf{x} al vector aleatorio normal de m dimensiones, con $E(\mathbf{x}) = \boldsymbol{\mu}$, donde $\boldsymbol{\mu}$ corresponde al vector media.

Supondremos que \mathbf{x} tiene una matriz V de variancia-covariancia dada por

$$V = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{bmatrix}. \quad (4-93)$$

En la expresión para V , convencionalmente σ_{ii} denota la variancia de la i -ésima componente, y σ_{ij} denota la covariancia entre las componentes i -ésima y j -ésima del vector aleatorio. La teoría nos asegura que V es una matriz siempre simétrica para la cual existe su inversa V^{-1} [2].

La función de densidad de probabilidad de \mathbf{x} está dada por:

$$f(\mathbf{x}) = |2\pi V|^{-1/2} \exp[-1/2(\mathbf{x} - \boldsymbol{\mu})'V^{-1}(\mathbf{x} - \boldsymbol{\mu})], \quad (4-94)$$

donde $|2\pi V|$ representa al determinante de la matriz $2\pi V$.

La integral de la ecuación (4-94) resulta muy complicada y desconocemos si existen tablas de áreas de probabilidad para vectores aleatorios normales con más de tres componentes [12]. Esta es una de las razones por las que podemos eventualmente necesitar el proceso de simulación de vectores normales al azar. La otra razón está dada por la frecuencia con que ocurre la dependencia entre variables normales en sistemas interdependientes.

En la generación de vectores normales al azar con un vector media y una matriz de variancia-covariancia dados, se requiere el empleo de un teorema [2, p. 19] que establece: si \mathbf{z} es un vector normal estándar, esto es, contiene como componentes las variables normales con media nula y variancia unitaria, entonces existe una única matriz bajo triangular C , tal que

$$\mathbf{x} = C\mathbf{z} + \boldsymbol{\mu}. \quad (4-95)$$

En este caso $(\mathbf{x} - \boldsymbol{\mu})$ tiene una matriz de variancia-covariancia dada por:

$$V = C \cdot C'. \quad (4-96)$$

A fin de obtener C de V se puede utilizar el llamado *método de la raíz cuadrada*, que proporciona un conjunto de fórmulas necesarias para la computación de los elementos de C [26].

$$\begin{aligned} c_{i1} &= \frac{\sigma_{i1}}{\sigma_{11}^{1/2}}, & 1 \leq i \leq m \\ c_{ii} &= \left(\sigma_{ii} - \sum_{k=1}^{i-1} c_{ik}^2 \right)^{1/2}, & 1 < i \leq m \\ c_{ij} &= \frac{\left(\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik}c_{jk} \right)}{c_{jj}}, & 1 < j < i \leq m. \end{aligned} \quad (4-97)$$

Puesto que C es bajo triangular, $c_{ij} = 0$ para toda $j > i$. Una vez obtenidos todos los elementos de C , se pueden determinar todas las componentes de x mediante sumas pesadas a partir de z :

$$x_i = \sum c_{ij} z_j + \mu_i. \quad (4-98)$$

La generación de un vector aleatorio x con media μ y una matriz de variancia-covariancia V , puede programarse de la siguiente manera.

1. Obtener a partir de V una matriz triangular C de acuerdo con la ecuación (4-97).
2. Generar m valores de variable aleatoria independientes con distribución normal estándar, cuya media sea cero y su variancia igual a uno, según la figura 4-16, o por medio de algún otro método equivalente.
3. Multiplicar y sumar los vectores y matrices tal como lo indican las expresiones (4-95) ó (4-98); en el resultado se obtendrá un vector aleatorio normal a partir de la distribución multivariada definida por μ , V y la ecuación (4-94).

Las técnicas para generar vectores tomando como base distribuciones normales multivariadas, se pueden emplear en aquellos problemas en los que se deben generar valores de variable aleatoria normales con una correlación previamente prescrita. Aquí se demuestra cómo funciona este orden de ideas cuando se tiene un caso con dos variables aleatorias normales x_1 y x_2 que están correlacionadas y para las cuales $E(x_1) = \mu_1$, $E(x_2) = \mu_2$, $\text{Var}(x_1) = \sigma_1^2$, $\text{Var}(x_2) = \sigma_2^2$, $\text{Cov}(x_1, x_2) = \rho\sigma_1\sigma_2$. En consecuencia,

$$V = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad (4-99)$$

$$C = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2 \sqrt{1 - \rho^2} \end{bmatrix}, \quad (4-100)$$

y

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Cz + \mu = \begin{bmatrix} \sigma_1 z_1 \\ \sigma_2(\rho z_1 + \sqrt{1 - \rho^2} \cdot z_2) \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}. \quad (4-101)$$

Esto significa que x_1 y x_2 se han generado como valores de variable aleatoria normales, que están correlacionadas con un coeficiente de correlación ρ . El ejemplo muestra cómo se pueden generar dos valores normales correlacionados partiendo de dos valores normales estándar independientes por medio del proceso de transformación descrito en la ecuación (4-95).

La distribución logarítmica normal

Si el logaritmo de una variable aleatoria tiene una distribución normal entonces la variable aleatoria tendrá una distribución continua sesgada positivamente, conocida con el nombre de distribución logarítmica normal. Frecuentemente se hace uso de esta distribución para describir procesos aleatorios que representan al producto de varios eventos pequeños e independientes [5]. Esta propiedad de la distribución logarítmica normal es conocida como *ley de efectos proporcionales*, que establecen una base sobre la que podemos decidir la validez con la cual una distribución logarítmica normal sirva o no para describir una variable aleatoria particular.

Las aplicaciones más importantes de la distribución logarítmica normal quedan comprendidas en áreas del análisis de rendimientos y ganancias, análisis de ventas y teoría de puntos de ruptura. Esta última proporciona criterios básicos para identificar la distribución de ciertos tamaños de lotes (inclusive tamaño de empresas o receptores de petróleo, etc.), así como también la distribución del ingreso, la cual se comporta en forma logarítmica normal [1, 8]. Por ejemplo, en lugar de estratificar un conjunto de datos numéricos asociados a algún concepto como el ingreso (o las ventas), utilizando una distribución normal, podemos emplear esta misma distribución sólo que aplicada a los logaritmos de los valores numéricos que queremos jerarquizar. En consecuencia, muy próximos al extremo superior de la escala logarítmica se encuentran pequeñas diferencias logarítmicas asociadas a las diferencias encontradas en los argumentos (los ingresos), mientras que en el extremo opuesto (de pequeños valores), las mismas diferencias en argumentos están asociadas con logaritmos cuya diferencia entre valores resulta mayor. Por consiguiente, la escala logarítmica proporciona el efecto de comprimir la distribución del ingreso en los niveles altos y ampliar la distribución en los niveles bajos. Una distribución de este tipo puede resultar con gran tendencia a cambiar, cualquier distribución sesgada positivamente en una distribución aproximadamente simétrica, lo cual constituye una de sus más valiosas características.

Se dice que X tiene una distribución logarítmica normal si cuando solamente se consideran los valores positivos de x , el logaritmo (en base e) de la variable aleatoria X tiene una función de densidad $f(y)$ dada como sigue:

$$f(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{y - \mu_y}{\sigma_y} \right)^2 \right] \quad -\infty < y < \infty, \quad (4-102)$$

con $y = \log x$. Los parámetros μ y σ_y^2 que aparecen en la expresión, corresponden a la media y variancia de y , respectivamente.

El valor esperado y la variancia de los valores de variable aleatoria, distribuidos en forma logarítmica normal x , están dados por las fórmulas siguientes:

$$EX = \exp\left(\mu_y + \frac{\sigma_y^2}{2}\right) \quad (4-103)$$

$$\begin{aligned} VX &= [\exp(2\mu_y + \sigma_y^2)][\exp(\sigma_y^2) - 1] \\ &= (EX)^2[\exp(\sigma_y^2) - 1]. \end{aligned} \quad (4-104)$$

La simulación de valores de variable aleatoria logarítmica normal con una media y variancia dadas, requiere necesariamente que μ_y y σ_y^2 estén expresadas en términos de EX y de VX , lo cual se puede lograr con sólo resolver la ecuación (4-104) para $\exp(\sigma_y^2)$. Tenemos, que

$$\frac{VX}{(EX)^2} = \exp(\sigma_y^2) - 1 \quad (4-105)$$

$$\exp(\sigma_y^2) = \frac{VX}{(EX)^2} + 1. \quad (4-106)$$

Tomando ahora el logaritmo de ambos miembros de la ecuación (4-106) obtenemos

$$\sigma_y^2 = \log\left[\frac{VX}{(EX)^2} + 1\right]. \quad (4-107)$$

A continuación, tomamos el logaritmo de ambos miembros de la ecuación (4-103)

$$\log(EX) = \mu_y + \frac{\sigma_y^2}{2} \quad (4-108)$$

y resolvemos para μ_y

$$\mu_y = \log(EX) - \frac{1}{2} \log\left[\frac{VX}{(EX)^2} + 1\right] \quad (4-109)$$

Ahora que tanto μ_y como σ_y^2 han quedado expresadas en términos de la media y la variancia de x , el valor logarítmico normal de la variable aleatoria que se va a generar o sea el valor de variable aleatoria z con distribución normal estándar, se puede definir como sigue:

$$z = \frac{\log x - \mu_y}{\sigma_y} \quad (4-110)$$

Resolviendo la ecuación (4-110) para $\log x$ y tomando el antilogaritmo de ambas partes de la ecuación, obtendremos

$$\log x = \mu_y + \sigma_y z \quad (4-111)$$

$$x = \exp(\mu_y + \sigma_y z). \quad (4-112)$$

Substituyendo el valor de z de la ecuación (4-73) en la ecuación (4-112), tendremos

$$x = \exp\left[\mu_y + \sigma_y\left(\frac{K}{12}\right)^{-1/2}\left(\sum_{i=1}^K r_i - \frac{K}{2}\right)\right] \quad (4-113)$$

o bien para los propósitos del programa en lenguaje FORTRAN, cuando $K = 12$,

$$X = \text{EXP}(EY + \text{STDY} * (\text{SUMR} - 6.0)). \quad (4-114)$$

Resumiendo, para generar valores logarítmicos normales de variable aleatoria x_1, x_2, \dots, x_n con EX y VX dadas, debemos en primer lugar determinar μ_y y σ_y de las ecuaciones (4-107) y (4-109), y después substituir estos valores en las ecuaciones (4-113) ó (4-114). Una vez definidos los valores de EY y de $STDY$ el procedimiento para simular valores logarítmicos normales diferirá del procedimiento para simular valores normales en mínimos detalles, considerando tan sólo que el reemplazo de la ecuación (4-114) por la proposición 6 de la subrutina FORTRAN que aparece en la figura 4-16, bastará para lograr los resultados deseados.

DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

Se encuentra definido un número muy significativo de distribuciones de probabilidad para variables aleatorias que solamente toman valores discretos, esto es, enteros no negativos. La distribución acumulativa de probabilidad para una variable aleatoria discreta X se define de manera muy similar a la de la ecuación (4-1).

$$F(x) = P(X \leq x) = \sum_{x=0}^x f(x), \quad (4-115)$$

donde $f(x)$ es la frecuencia o función de probabilidad de X , definida por valores enteros x tales que:

$$f(x) = P(X = x) \quad (4-116)$$

para $x = 0, 1, 2, \dots$

Las distribuciones discretas de probabilidad son muy útiles cuando se las emplea como modelos estocásticos para ciertos procesos de *conteo*, ya sea sobre muestras finitas o no finitas, donde la presencia o ausencia de un atributo dicotómico está gobernada por el azar. Desde un punto de vista empírico las distribuciones discretas pueden también ocurrir como resultado de redondear medidas continuas sobre una escala discreta. Sin embargo, estrictamente hablando, las distribuciones discretas de probabilidad resultan ser los modelos más apropiados para fenómenos aleatorios cuando

los valores de las variables aleatorias se pueden determinar por medio de procesos de conteo.

Las secciones siguientes contienen la descripción de técnicas para la generación de valores de variables estocásticas a partir de la mayoría de las distribuciones discretas de probabilidad más conocidas. El formato empleado en la presentación de estos métodos es semejante a la forma en que presentamos las distribuciones continuas.

La distribución geométrica

Entre las primeras y probablemente más simples de las formulaciones matemáticas de procesos estocásticos, se encuentra la llamada *de ensayos de Bernoulli*. Estos ensayos son experimentos independientes al azar, en los que el resultado de cada ensayo queda registrado, ya sea como un éxito o un fracaso. La probabilidad de éxito se denota por p ($0 \leq p \leq 1$) y se supone que p es constante para cualquier sucesión particular de ensayos. La probabilidad de un fracaso se denota por q , donde

$$q = 1 - p. \quad (4-117)$$

Una sucesión de ensayos Bernoulli, combinada con cierto proceso de conteo, viene a constituir la base conceptual para una gran familia de distribuciones discretas de probabilidad, incluyendo la geométrica, binomial negativa, Poisson y otras distribuciones binomiales. Los valores de variable aleatoria que se generan al contar el número de fracasos en una sucesión de ensayos (o eventos) antes de que ocurra el primer éxito, son valores de variable aleatoria que se ajustan a una distribución geométrica. La distribución geométrica de probabilidad tiene un gran valor y utilidad en el área del control estadístico de calidad, así como también para las distribuciones de rezagos y movimientos en modelos econométricos.

La distribución geométrica queda descrita por la siguiente función de probabilidad:

$$f(x) = pq^x \quad x = 0, 1, 2, \dots \quad (4-118)$$

y la función de distribución acumulativa está definida por

$$F(x) = \sum_{x=0}^{\infty} pq^x \quad X = 0, 1, 2, \dots, x. \quad (4-119)$$

Puesto que por definición se tiene $F(x) = P(X \leq x)$, y como $P(X = 0) = F(0) = p$, el rango de $F(x)$ es $p \leq F(x) \leq 1$. Por otra parte, $P(X > x) = 1 - F(x)$, lo que implica que $P(X > 0) = q$ y además

$$1 - F(x) = q^{x+1}. \quad (4-120)$$

El valor esperado y la variancia de la variable geométrica, están dados por

$$EX = \frac{q}{p} \quad (4-121)$$

$$VX = \frac{q}{p^2} = \frac{EX}{p}. \quad (4-122)$$

Esta última expresión implica que la variancia difiere de la media en un factor de $1/p$. De la ecuación (4-118) resulta claro cómo la distribución geométrica siempre tendrá el perfil de una J con modo en el punto $x = 0$ [23].

La distribución geométrica tiene sólo un parámetro p , el cual se puede expresar como una función de la media EX

$$p = \frac{1}{1 + EX}. \quad (4-123)$$

Para generar en una computadora valores de variable aleatoria con distribución geométrica se emplea la técnica de la transformación inversa y la fórmula que aparece en la ecuación (4-120) [11]. Al observar que el rango de la expresión $[1 - F(x)/q]$ es unitario, resulta que

$$r = q^x, \quad (4-124)$$

y consecuentemente

$$x = \frac{\log r}{\log q}, \quad (4-125)$$

donde al valor x siempre se le redondea al entero que sea menor. Esto se puede lograr de manera muy simple con sólo retener los números hasta antes del punto decimal, o bien convirtiendo los números en modo de punto flotante al de punto fijo [15]. Para generar un valor de variable aleatoria con distribución geométrica haciendo uso de esta técnica, se requiere únicamente el empleo de un número aleatorio uniforme, tal como lo muestra la ecuación (4-125).

En las figuras 4-17 y 4-18 de la siguiente sección, se describe un diagrama de flujo y una subrutina en FORTRAN que sirven para el proceso de generación de valores de variable aleatoria con distribución geométrica cuando se tiene una probabilidad fija dada, para fracasos q y para los éxitos k igual a uno.

Se tiene un método alternativo que utiliza la técnica de rechazo para generar valores con distribución geométrica, capaces de reproducir los ensayos de Bernoulli en una computadora. Por lo común, se prefiere este último método, en relación al que se menciona en primer término, cuando

se requiere una mejor precisión para grandes valores de p . En primer lugar, definimos una variable x que se usará como contador, por lo que se la iguala a un valor nulo. A continuación generamos una sucesión de valores de variable aleatoria uniformes $r_1, r_2, \dots, r_i, \dots$, la cual termina cuando alcanzamos un valor de r_i que resulta ser menor o igual que p . Para cada valor de r_i en la sucesión, que sea mayor que p , incrementamos el valor de x en uno. En otras palabras, contaremos el número de fallas o sea el número de veces que r_i tiene un valor mayor que p . Cuando logramos el primer valor de r_i menor que p se termina la sucesión y el valor de x corresponde al valor de la variable aleatoria con distribución geométrica. Después de reinstaurar el valor de x a cero se genera una segunda sucesión que conduce a un segundo valor de x .

En la presentación anterior se definió a x como el número de fallas que ocurren antes del primer éxito; sin embargo, se puede volver a definir a x en forma tal que incluya, tanto el número de fracasos como al primer éxito. Pese a que el proceso para generar valores de variable aleatoria con distribución geométrica, es similar al procedimiento previamente explicado, las ecuaciones (4-118) y (4-121) se reformulan de la siguiente manera:

$$f(x) = pq^{x-1} \quad x = 1, 2, \dots \quad (4-126)$$

$$EX = \frac{1}{p}. \quad (4-127)$$

La distribución binomial negativa

Cuando los procesos de ensayos de Bernoulli, tal como se han descrito en la sección anterior, se repiten hasta lograr que ocurran k éxitos ($k > 1$), la variable aleatoria que caracteriza al número de fallas tendrá una distribución binomial negativa. Por consiguiente, los valores de variables aleatorias con distribución binomial negativa coinciden esencialmente con la suma de k valores de variable aleatoria con distribución geométrica; en este caso, k es un número entero y la distribución recibe el nombre de distribución de Pascal. En consecuencia, la distribución geométrica constituye un caso particular de la distribución de Pascal, especificada para k igual a uno.

La función de distribución de probabilidad para una distribución binomial negativa está dada por

$$f(x) = \binom{k+x-1}{x} p^k q^x \quad x = 0, 1, 2, \dots, \quad (4-128)$$

donde k es el número total de éxitos en una sucesión de $k+x$ ensayos, con x el número de fallas que ocurren antes de obtener k éxitos. El valor

esperado y la variancia de X se representa con:

$$EX = \frac{kq}{p} \quad (4-129)$$

$$VX = \frac{kq}{p^2}. \quad (4-130)$$

Se debe hacer notar que tanto la distribución geométrica como la binomial negativa se caracterizan por una sobredispersión, esto es, $VX > EX$.

Para una media y una variancia dadas, se pueden determinar los parámetros p y k de la siguiente manera

$$p = \frac{EX}{VX} \quad (4-131)$$

$$k = \frac{(EX)^2}{VX - EX}. \quad (4-132)$$

Sin embargo, puede suceder que el proceso de simulación se complique considerablemente cuando resulte que en la ecuación (4-132) el valor que se obtenga al efectuar el cómputo de k no sea un entero.

Cuando k es un entero, los valores de la variable aleatoria con distribución de Pascal se pueden generar con sólo considerar la suma de k valores con distribución geométrica. En consecuencia:

$$x = \frac{\left(\sum_{i=1}^k \log r_i\right)}{\log q} = \frac{\log \left(\prod_{i=1}^k r_i\right)}{\log q} \quad (4-133)$$

viene a ser un valor de variable aleatoria con distribución de Pascal, una vez que su magnitud se redondea con respecto al menor entero más próximo al valor calculado.

El contenido de las figuras 4-17 y 4-18 representa, respectivamente, al diagrama de flujo y a la subrutina FORTRAN para generar los valores que siguen la distribución de Pascal, empleando el método ya mencionado. X es una variable FORTRAN de tipo entero que corresponde a x . Obsérvese que, en la séptima proposición FORTRAN, automáticamente tiene lugar el proceso de redondeo de acuerdo con la aritmética entera FORTRAN.

Cuando k no resulte ser un entero, deberemos confinarnos a los métodos de aproximación para generar valores de variable aleatoria con distribución binomial negativa. Uno de tales métodos implica la generación de una mezcla de valores de variable aleatoria con dos valores enteros, diferentes para k . Por ejemplo, si k es igual a 3.60 podríamos generar una

mezcla de valores de variable aleatoria con distribución de Pascal: k igual a 3 en un caso y con k igual a 4 en otro; aunque tomando siempre en cuenta que el valor esperado de k es igual a 3.60. Otra alternativa es la que considera la generación de valores de variable aleatoria con distribución de Poisson, cuyo único parámetro se ajuste a una distribución gamma de parámetros k y α [10]. Los valores binomiales negativos que se generan en este último criterio tendrán los parámetros k y p , en donde

$$p = \frac{\alpha}{1 + \alpha} \quad (4-134)$$

Los valores de variable aleatoria con distribución de Pascal también se pueden definir como el número total de ensayos, de tal suerte que ocurren x fallas antes de que ocurran k éxitos. En este caso, el valor de la variable aleatoria ajustado a la distribución de Pascal estará dado por $k + x$ [6]. De acuerdo con esta proposición, la media de estos valores está definida por

$$E(k + x) = \frac{kq}{p} + k = \frac{k}{p} \quad (4-135)$$

La distribución binomial

Las variables aleatorias definidas por el número de eventos exitosos en una sucesión de n ensayos independientes de Bernoulli, para los cuales la probabilidad de éxito es p en cada ensayo, siguen una distribución binomial. Este modelo estocástico también se puede aplicar al proceso de muestreo aleatorio con reemplazo, cuando los elementos muestreados tienen sólo dos tipos de atributos (por ejemplo *si* y *no*, o respuestas como *defectuoso* o *aceptable*). El diseño de una muestra aleatoria de n elementos es análoga a n ensayos independientes de Bernoulli, en los que x es un valor binomial que está

denotando al número de elementos de una muestra de tamaño n con atributos idénticos. Es ésta la analogía que sitúa la distribución binomial como uno de los modelos más importantes en las áreas del muestreo estadístico y del control de calidad.

1. SUBROUTINE PASCAL (K, Q, X)
2. TR = 1.0
3. QR = LOG (Q)
4. DO 6 I = 1, K
5. R = RND(R)
6. TR = TR * R
7. NX = LOG (TR)/QR
8. X = NX
9. RETURN

Figura 4-18. Subrutina FORTRAN para la generación de valores de variable aleatoria con distribución de Pascal.

La distribución binomial proporciona la probabilidad de que un evento o acontecimiento tenga lugar x veces en un conjunto de n ensayos, donde la probabilidad de éxito está dada por p . La función de probabilidad para la distribución binomial se puede expresar de la manera siguiente:

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad (4-136)$$

donde x se toma como un entero definido en el intervalo finito $0, 1, 2, \dots, n$, y al que se le asocia el valor $q = (1 - p)$.

El valor esperado y la variancia de la variable binomial X son

$$EX = np \quad (4-137)$$

$$VX = npq. \quad (4-138)$$

La segunda expresión implica que la variancia de las variables binomiales siempre tiene un valor menor al de la media. Aún más, nótese cómo se define en la ecuación (4-136) la distribución de $(n - x)$ con un valor esperado correspondiente de nq .

Cuando se conocen la media y la variancia, resulta inmediata la determinación de p y de n , las cuales pueden calcularse como sigue:

$$p = \frac{(EX - VX)}{EX} \quad (4-139)$$

$$n = \frac{(EX)^2}{(EX - VX)}. \quad (4-140)$$

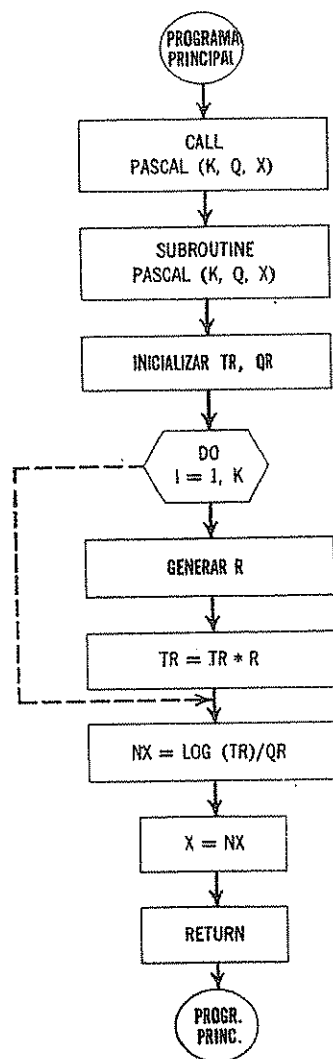


Figura 4-17. Diagrama de flujo, empleado para generar valores de variable aleatoria con una distribución de Pascal.

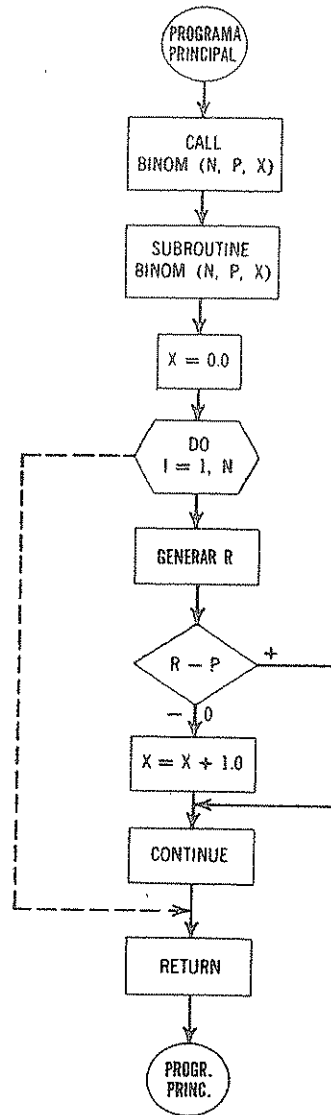


Figura 4-19. Diagrama de flujo para la generación de valores de variable aleatoria con distribución binomial.

Un segundo método para generar valores binomiales es el que se basa en las sumas aleatorias de valores de variables aleatorias con distribución

La distribución normal proporciona, cuando n es muy grande, una buena aproximación para la distribución binomial. Puesto que con la distribución normal resulta posible manipular valores negativos, a fin de hacer un buen uso de tal distribución la probabilidad de registrar observaciones negativas deberá ser despreciablemente pequeña. En la práctica esto significa que el valor esperado deberá ser por lo menos tres veces mayor que la desviación estándar, o sea

$$np \geq 3(npq)^{1/2} \quad (4-141)$$

lo cual implica que $n \geq 9q/p$.

Los valores de variable aleatoria con distribución binomial se pueden generar de muy diversos modos, aunque uno de los métodos más simples, que en el caso de que el valor de n sea moderado resulta uno de los métodos más eficientes, es el basado en la reproducción de ensayos de Bernoulli, siguiendo el método de rechazos. Este método empieza con valores conocidos de p y de n y consiste en generar n números aleatorios después de fijar x_0 igual a cero. Para cada número aleatorio r_i ($1 \leq i \leq n$) se efectúa una prueba y la variable x_i se incrementa de acuerdo con el siguiente criterio:

$$x_i = x_{i-1} + 1 \quad \text{si } r_i \leq p \quad (4-142)$$

$$x_i = x_{i-1} \quad \text{si } r_i > p. \quad (4-143)$$

Después de haberse generado n números aleatorios, el valor de x_n será igual al valor de la variable aleatoria con distribución binomial x . Este procedimiento se puede repetir tantas veces como valores binomiales se requieran.

1. SUBROUTINE BINOM (N, P, X)
2. X = 0.0
3. DO 7 I = 1, N
4. R = RND (R)
5. IF (R - P) 6, 6, 7
6. X = X + 1.0
7. CONTINUE
8. RETURN

Figura 4-20. Subrutina FORTRAN para generar valores de variable aleatoria con distribución binomial.

geométrica, con lo cual se obtiene el número de éxitos en n ensayos. En el caso de que p sea pequeño, este método puede ser mucho más rápido.

Las figuras 4-19 y 4-20 muestran, respectivamente, el diagrama de flujo y la subrutina FORTRAN para generar los valores binomiales, siguiendo el primero de los dos métodos presentados.

La distribución hipergeométrica

Considérese una población que consta de N elementos tales que cada uno de ellos pertenece a la clase I o a la II. Denotemos por Np al número de elementos que pertenecen a la clase I y por Nq al número de elementos que son miembros de la clase II, donde $p + q = 1$. Si en una población de N elementos se toma una muestra aleatoria que conste de n elementos ($n < N$) sin que tenga lugar algún reemplazo, entonces el número de elementos x de la clase I en la muestra de n elementos, tendrá una distribución de probabilidad hipergeométrica [31, p. 133]. En las áreas de control de calidad y en la de control de producción, con mayor frecuencia se encuentran las aplicaciones de la distribución hipergeométrica. Por ejemplo, si el intervalo entre la llegada de órdenes sucesivas de los clientes relativas a cierto producto de la compañía, se distribuye geoméricamente, entonces la demanda total en cualquier período dado de tiempo tendrá una distribución hipergeométrica [5, p. 166].

La distribución hipergeométrica está descrita por la siguiente función de probabilidad:

$$f(x) = \frac{\binom{Np}{x} \binom{Nq}{n-x}}{\binom{N}{n}} \quad \begin{matrix} 0 \leq x \leq Np \\ 0 \leq n-x \leq Nq, \end{matrix} \quad (4-144)$$

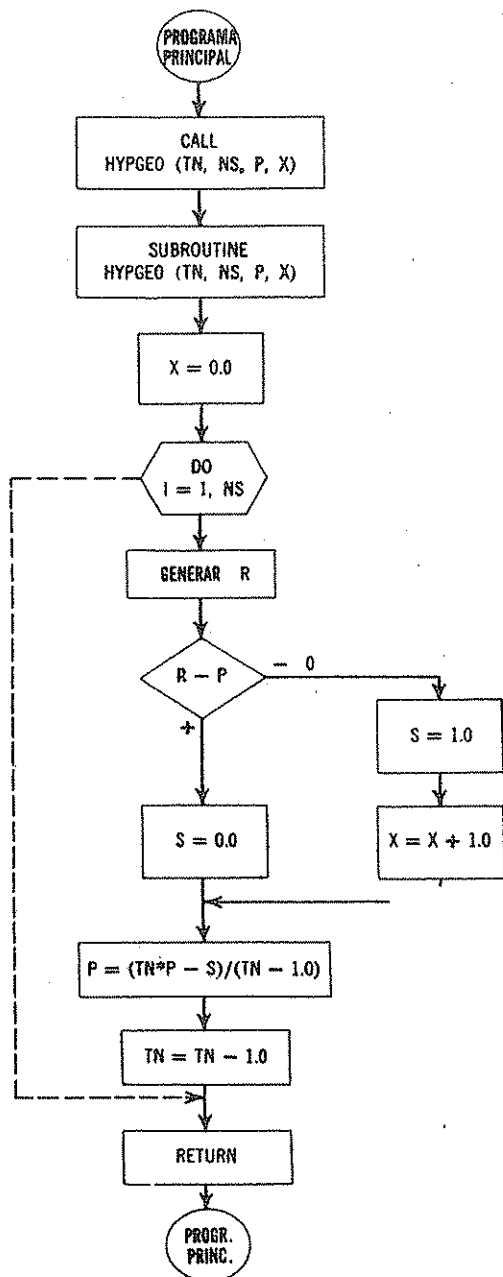


Figura 4-21. Diagrama de flujo para la generación de valores de variable aleatoria con distribución hipergeométrica.

donde x , n y N son enteros. El valor esperado y la variancia se caracterizan como sigue:

$$EX = np \quad (4-145)$$

$$VX = npq \left(\frac{N-n}{N-1} \right) \quad (4-146)$$

La generación de valores hipergeométricos involucra, substancialmente, la simulación de experimentos de muestreo *sin reemplazo*. En otras palabras, bastará sencillamente con que alteremos el método de ensayos de

1. SUBROUTINE HYPGEO (TN, NS, P, X)
2. X = 0.0
3. DO 11 I = 1, NS
4. R = RND (R)
5. IF (R - P) 6, 6, 9
6. S = 1.0
7. X = X + 1.0
8. GO TO 10
9. S = 0.0
10. P = (TN * P - S) / (TN - 1.0)
11. TN = TN - 1.0
12. RETURN

Figura 4-22. Subrutina FORTRAN para la generación de valores de variable aleatoria con distribución hipergeométrica.

Barnoulli para generar valores binomiales, con objeto que N y p varíen en forma dependiente respecto al número total de elementos que previamente se han obtenido entre la población y el número de elementos de la clase I que se han extraído. A medida que se extrae un elemento de una muestra de n elementos, se reduce el valor de $N = N_0$ de acuerdo con la fórmula:

$$N_i = N_{i-1} - 1 \quad i = 1, 2, \dots, n. \quad (4-147)$$

De manera similar, el valor de $p = p_0$ se transforma según

$$p_i = \frac{N_{i-1}p_{i-1} - S}{N_{i-1} - 1} \quad i = 1, 2, \dots, n, \quad (4-148)$$

a medida que se saca el i -ésimo elemento de la muestra de n elementos, donde $S = 1$ cuando el elemento de muestra $(i-1)$ pertenece a la clase I y $S = 0$ cuando el elemento de muestra $(i-1)$ pertenece a la clase II. Ciertamente, los valores iniciales de N_0 y P_0 corresponden: a N el tamaño

inicial de la población y a p la proporción de la población total que consta de elementos de la clase I.

El diagrama de flujo y la subrutina FORTRAN para generar los valores hipergeométricos se describen en las figuras 4-21 y 4-22, respectivamente. Los símbolos TN y NS se han empleado en este caso para denotar, respectivamente, a N y n .

La distribución de Poisson

Si tomamos una serie de n ensayos independientes de Bernoulli, en cada uno de los cuales se tenga una probabilidad p muy pequeña relativa a la ocurrencia de un cierto evento, a medida que n tiende al infinito, la probabilidad de x ocurrencias está dada por la distribución de Poisson

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots \quad (4-149)$$

$$\lambda > 0,$$

siempre y cuando permitamos que p se aproxime a cero de manera que se satisfaga la relación $\lambda = np$ consistentemente. De nuestra discusión previa sabemos que np es el valor esperado de la distribución binomial y se puede demostrar que λ es el valor esperado para la distribución de Poisson. De hecho, tanto el valor esperado como la variancia de la distribución de Poisson coinciden en el valor λ . También se puede demostrar que si x es una variable de Poisson con parámetro λ , entonces para valores muy grandes de λ ($\lambda > 10$), se puede utilizar la distribución normal con $EX = \lambda$ y $VX = \lambda$ para aproximar la distribución de x .

Los eventos que se distribuyen en forma poissoniana ocurren frecuentemente en la naturaleza; por ejemplo, el número de aviones que descienden en un aeropuerto en un período de veinticuatro horas puede ser considerablemente grande. Aun así, resulta muy pequeña la probabilidad de que un avión aterrice durante un segundo determinado. Por lo tanto, podemos esperar que en un período determinado, la probabilidad de que desciendan 0, 1, 2, ... aviones, obedecerá a las leyes de la distribución de Poisson. Esta distribución es particularmente útil cuando tratamos con problemas en los que se da la ocurrencia de eventos aislados sobre un intervalo continuo de tiempo, o bien cuando resulta posible prescribir el número de veces que ocurre un evento aunque no el número de veces que no ocurre.

Para simular una distribución de Poisson con parámetro λ , nos podemos servir ventajosamente de la relación conocida entre las distribuciones exponenciales y de Poisson. Se puede justificar que si 1) el número total de eventos que ocurren durante un intervalo de tiempo dado es inde-

pendiente del número de eventos que ya han ocurrido previamente al inicio del intervalo y 2) la probabilidad de que un evento ocurra en el intervalo de t a $t + \Delta t$ es aproximadamente $\lambda \Delta t$ para todos los valores de t , entonces: a), la función de densidad del intervalo t entre las ocurrencias de eventos consecutivos es $f(t) = \lambda e^{-\lambda t}$, y b), la probabilidad de que ocurran x eventos durante el tiempo t es

$$f(x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!} \quad (4-150)$$

para toda x y toda t .

Considérese un horizonte de tiempos que se inicia en el origen 0 como punto de referencia y que se ha dividido en intervalos unitarios, como se ilustra en la figura 4-23. Supóngase también que los eventos ocurren a lo largo del mencionado horizonte y que se denotan por el símbolo (\wedge) . Se supondrá que el intervalo t entre eventos obedece a una distribución exponencial

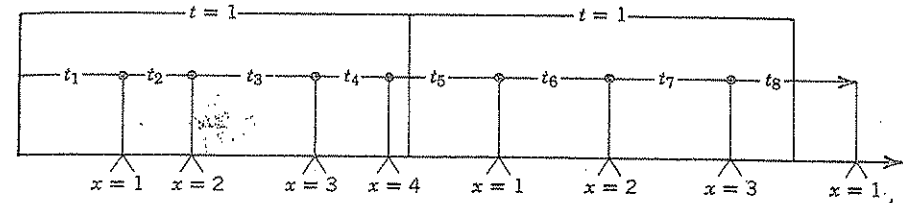


Figura 4-23. Eventos distribuidos en forma poissoniana, sobre una escala de tiempo.

cuyo valor esperado es igual a $1/\lambda$. Con estos antecedentes se implica que el número de eventos x que ocurren durante un tiempo unitario seguirán una distribución de Poisson cuyo valor esperado estará dada por λ . Un método para generar valores de variable aleatoria con distribución de Poisson deberá considerar la generación de intervalos t_1, t_2, t_3, \dots , distribuidos en forma exponencial con un valor esperado igual a 1. Un vez generados estos intervalos aleatorios, se acumulan hasta que su suma exceda al valor de λ .

En términos matemáticos el valor poissoniano x se determina haciendo uso de la siguiente desigualdad:

$$\sum_{i=0}^x t_i \leq \lambda < \sum_{i=0}^{x+1} t_i \quad (x = 0, 1, 2, \dots), \quad (4-151)$$

donde los valores de la variable aleatoria t_i se generan por medio de la fórmula

$$t_i = -\log r_i \quad (4-152)$$

con una media unitaria. Un método más rápido para generar los valores

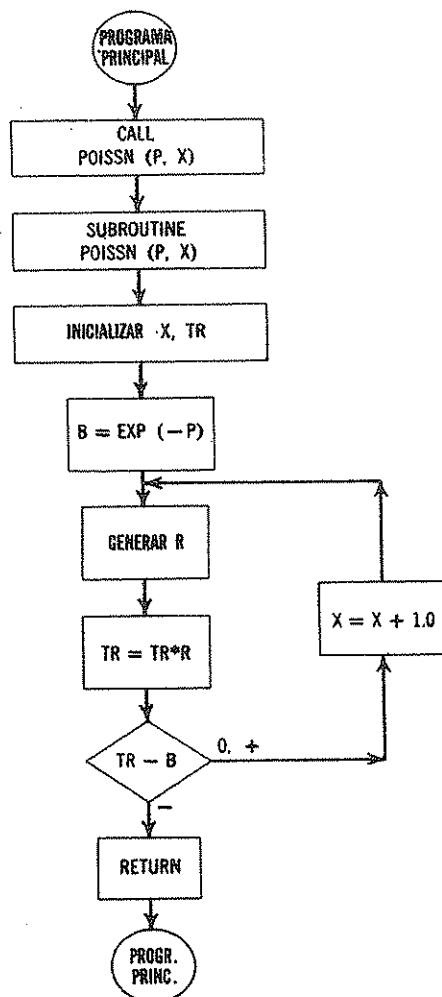


Figura 4-24. Diagrama de flujo para la generación de valores de variable aleatoria con distribución de Poisson.

poissonianos x [28, p. 37] es el que consiste en reformular la ecuación (4-151) de la manera siguiente:

$$\prod_{i=0}^x r_i \geq e^{-\lambda} > \prod_{i=0}^{x+1} r_i. \quad (4-153)$$

La subrutina FORTRAN correspondiente a la ecuación (4-153) aparece en las figuras 4-24 y 4-25, donde P representa la constante FORTRAN que corresponde al parámetro λ .

1. SUBROUTINE POISSN (P, X)
2. X = 0.0
3. B = EXP (-P)
4. TR = 1.0
5. R = RND (R)
6. TR = TR * R
7. IF (TR - B) 10, 8, 8
8. X = X + 1.0
9. GO TO 5
10. RETURN

Figura 4-25. Subrutina FORTRAN para generar valores de variable aleatoria que obedecen la distribución de Poisson.

Distribuciones discretas empíricas

En este capítulo nos hemos concentrado en los métodos para generar ciertas distribuciones particulares de probabilidad, como la normal, la binomial y la de Poisson, para mencionar sólo algunas. Ahora trataremos un método un tanto más general que puede emplearse para simular cualquiera de las siguientes distribuciones: 1) empírica, 2) discreta y 3) continua que pueda ser aproximada mediante una distribución discreta. Sin embargo, en general nos abstendremos de utilizar este método para generar valores de variable aleatoria a partir de distribuciones de probabilidad estándares, debido a que cualquiera de los métodos descritos previamente se espera que proporcione resultados más satisfactorios, desde el punto de vista de la velocidad de computación, facilidad de programación y requisitos de almacén o memoria. En otros términos, el método que se propondrá en esta sección es uno que se utiliza siempre y cuando no se disponga de otra alternativa.

Sea X una variable aleatoria discreta con $P(X = b_i) = p_i$, tal como se presenta la variable aleatoria de la siguiente tabla:

| b_i | $P(x = b_i) = p_i$ |
|----------|--------------------|
| b_1 | 0.273 |
| b_2 | 0.037 |
| b_3 | 0.195 |
| b_4 | 0.009 |
| b_5 | 0.124 |
| b_6 | 0.058 |
| b_7 | 0.062 |
| b_8 | 0.151 |
| b_9 | 0.047 |
| b_{10} | 0.044 |

En consecuencia, resulta evidente que un método para generar x en una computadora es aquél que genera un valor de variable aleatoria r sujeto a una distribución uniforme, en el intervalo $(0, 1)$ y un conjunto de valores $x = b_i$ siempre que se satisfaga

$$p_1 + \dots + p_{i-1} < r \leq p_1 + \dots + p_i. \quad (4-154)$$

Pese a que se han desarrollado un buen número de técnicas de búsqueda basadas en este método, en su gran mayoría requieren programas relativamente complejos que a su vez emplean un tiempo de computación excesivo.

Uno de los procedimientos más rápidos para generar valores de variable aleatoria discretos es el desarrollado por G. Marsaglia [17], quien presupone la disponibilidad de una computadora decimal cuyos bloques o palabras de memoria pueden referirse mediante números. Esta última característica, en realidad, constituye una propiedad de la gran mayoría de las computadoras actuales. En el procedimiento de Marsaglia se mencionan los números 273 b_1 , 37 b_2 , 195 b_3 , ..., 44 b_{10} y los que pertenecen a cada una de estas clases respectivas, en las ubicaciones de memoria desde la 0 a la 999. Entonces, si por ejemplo $r = .d_1d_2d_3d_4$ es un número aleatorio uniforme de cuatro dígitos, generado por la computadora, el número en la localidad $d_1d_2d_3$ será el valor de x .

Conviene hacer notar que si bien este método es extremadamente rápido, también requiere por lo menos una memoria de 1000 palabras. Existe otro método desarrollado también por Marsaglia que en forma alternativa utiliza mucho menos capacidad de memoria, aunque incrementa ligeramente el tiempo de computación [17].

Cadenas discretas de Markov

Una posibilidad, muy frecuente por cierto, de caracterizar los sistemas operacionales radica en la especificación del sistema en términos de una sucesión de estados distinguibles: por ejemplo, el estado en que se encuentra cierto instrumental de producción, la cola o las facilidades de almacenamiento, pueden describirse adecuadamente mediante el número de objetos estacionados en cualquier tiempo dado. En caso de que el tiempo pueda medirse mediante unidades discretas, podremos describir los cambios que ocurren en los estados del sistema, de la siguiente manera. En cualquier momento en que el sistema se encuentre en el estado i al inicio de un período, se puede definir la probabilidad de que el sistema evolucione al estado j cuando se inicie el principio del siguiente período. Esta probabilidad p_{ij} dependerá tan sólo de los estados i, j y para cada i se tendrá que $\sum_j p_{ij} = 1$.

Las p_{ij} pueden disponerse en un arreglo matricial que recibe el nombre de matriz de probabilidad de transición $P = ||p_{ij}||$, la cual determina completamente al comportamiento del sistema. Esta clase de comportamientos

recibe el nombre de proceso de Markov y la sucesión de transiciones que sirve de muestra se la conoce por el nombre de cadena de Markov.

Cabe mencionar que aunque existen métodos analíticos para evaluar la distribución de probabilidad de los procesos de Markov bajo ciertas condiciones, también se puede recurrir a la determinación bajo estimaciones de muestreo basadas en la simulación de las cadenas de Markov. Esta técnica permite hacer las estimaciones de las distribuciones de frecuencia de los vectores de probabilidad, tanto de estados transitorios como estacionarios. Una alternativa más es la de considerar la posibilidad que existe para analizar las probabilidades de transición no constante, haciendo uso de la simulación.

Uno de los métodos conocidos [11, p. 237] para generar cadenas de Markov utiliza los renglones de la matriz de transición $||p_{ij}||$ de modo muy similar al descrito en la sección previa. Si el último estado del sistema fue i , entonces el próximo estado será j , si

$$\sum_{k=1}^{j-1} p_{ik} < r \leq \sum_{k=1}^j p_{ik}, \quad (4-155)$$

donde r es un número aleatorio uniforme en el intervalo $(0, 1)$. Cada número que se genere originará una transición del estado i al estado j y el programa podrá incluir la facilidad de un contador de la frecuencia en que el sistema se localiza en cada uno de los estados finitos.

En las figuras 4-26 y 4-27 se describe un programa FORTRAN que sirve para generar la distribución de frecuencias de los estados por medio de la simulación de una cadena de Markov. Este programa está dimensionado hasta una matriz de transición de 10×10 . En este caso M indica la dimensión real de la matriz y N representa la longitud deseada (el número de transiciones) de la cadena, e I denota al estado inicial seleccionado. Las declaraciones FORMAT que sirven de formato para la lectura

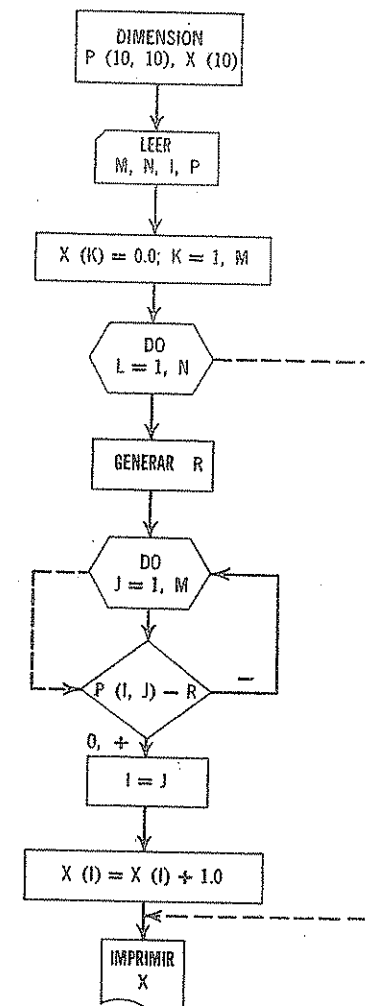


Figura 4-26. Diagrama de flujo para la simulación de cadenas de Markov.

de datos, se simbolizan sin otra especificación ulterior, tan sólo por paréntesis. La matriz P en el programa *no* es la matriz de probabilidad de transición sino que es una derivada de la matriz de transiciones, de modo tal que sólo contiene las probabilidades acumuladas de cada renglón. El vector X deberá contener, al final de una sucesión de N transiciones, la distribución de frecuencias de los estados. El programa puede genera-

```

1. DIMENSION P(10, 10), X(10)
2. READ ( ), M, N, I
3. READ ( ), P
4. DO 5 K = 1, M
5. X(K) = 0.0
6. DO 12 L = 1, N
7. R = RND (R)
8. DO 10 J = 1, M
9. IF (P(I,J) - R) 10, 11, 11
10. CONTINUE
11. I = J
12. X(I) = X(I) + 1.0
13. PRINT ( ), X
14. END

```

Figura 4-27. Programa FORTRAN para la simulación de cadenas de Markov.

lizarse con sólo añadir subrutinas que alimenten información reacondicionada de entrada o que efectúe determinadas computaciones sobre el vector de frecuencias X .

SERIES DE TIEMPO AUTOCORRELACIONADAS

Si en un proceso estocástico se producen variables aleatorias tales que se tenga una variable aleatoria x_t para cada valor de t , donde t representa al tiempo, entonces nos encontramos frente a una función aleatoria del tiempo llamada serie de tiempo. Cuando estas series de tiempo se relacionan específicamente con fenómenos económicos o tecnológicos, aparece una propiedad muy común en estas series, la cual establece que la covariancia de x_{t+k} y x_t , donde k es el *rezago* (i.e. el número de intervalos de tiempo entre los valores respectivos de las series de tiempo), resulta con un valor no negativo. Para el rezago K , definimos la función de covariancia $\phi(k)$ como sigue:

$$\phi(k) = E(x_t, x_{t+k}) \quad (4-156)$$

y la función de autocorrelación $\rho(k)$ como

$$\rho(k) = \frac{\phi(k)}{\phi(0)} \quad (4-157)$$

En ambas expresiones estamos suponiendo que $E(x_t) = E(x_{t+k}) = 0$ y que tanto $\phi(k)$ como $\rho(k)$ son funciones de k únicamente. Cuando es tal el caso, estas condiciones resultan válidas únicamente para series estacionarias de tiempo [31, p. 516].

No es posible, por lo común, generar series de tiempo con una función de autocorrelación arbitraria; sin embargo, se tienen dos funciones especiales que pueden usarse con una flexibilidad satisfactoria, si es que la distribución de las variables x_t es normal, de media nula y con una variancia de constante idéntica [5, p. 169].

Función lineal de autocorrelación

Sean

$$\rho(k) = 1 - \frac{k}{m} \quad k \leq m \quad (4-158)$$

$$\rho(k) = 0 \quad k > m$$

Estas expresiones representan una función linealmente decreciente en k a la vez que un modelo para las series de tiempo autocorrelacionadas, en las que puede suponerse una autocorrelación cero para rezagos mayores que m . La técnica para generar una serie de tiempo con esta función de autocorrelación se basa en el proceso para generar valores de variable aleatoria que obedecen a una distribución normal, como lo describe la ecuación (4-75) y suponiendo además, que los números aleatorios uniformes se transforman en valores cuya esperanza es nula, es decir, $E(r) = 0$. Entonces, si

$$x_t = \sum_{j=1}^N r_j \quad (4-159)$$

x_t tendrá media cero y variancia $N\sigma^2$ con $\sigma^2 = \text{Var}(r)$. El siguiente valor se genera con

$$x_{t+1} = \sum_{j=p+1}^{N+p} r_j \quad (4-160)$$

en donde $(N - p)$ de los r_j números aleatorios corresponden a términos comunes que aparecen en las sumas sucesivas.

La función de autocorrelación con rezago k se deriva de la siguiente relación de identidad:

$$(x_t - x_{t+k})^2 = x_t^2 - 2x_t x_{t+k} + x_{t+k}^2 \quad (4-161)$$

Regresando a los valores esperados y con ayuda de las ecuaciones (4-165) y (4-159) podemos escribir

$$E(x_t - x_{t+k})^2 = 2N\sigma^2 - 2\phi(k) \quad (4-162)$$

El paréntesis izquierdo contiene tan sólo $2kp$ números r_i no nulos e independientes, cuya variancia es igual a $2kp\sigma^2$. Consecuentemente, si $k \leq N/p$

$$kp\sigma^2 = N\sigma^2 - \phi(k) \quad (4-163)$$

y

$$\rho(k) = \frac{\phi(k)}{N\sigma^2} = 1 - \frac{kp}{N} \quad (4-164)$$

Se acostumbra elegir N igual a 12, con lo cual se logra obtener una variancia unitaria para los x_i . Sin embargo, en este caso la función de autocorrelación se define sólo para valores de k a lo más $k = 12/p$ rezagos. Obsérvese que la fórmula (4-164) corresponde a la ecuación (4-158) con $m = N/p$.

Las subrutinas FORTRAN escritas para generar una serie de tiempo autocorrelacionada de valores x distribuidos normalmente y con una media igual a cero y una variancia unitaria, quedan descritos en las figuras 4-28 y 4-29. El valor de p en la ecuación (4-160) corresponde con la variable K del programa FORTRAN. En esta subrutina se presuponen doce valores iniciales para los valores uniformes (RX) y se produce un valor X de la serie de tiempo después de cada proposición CALL.

Función exponencial de autocorrelación

El coeficiente de correlación con rezago k se expresa como

$$\rho(k) = \lambda^k; \quad 0 < \lambda < 1. \quad (4-165)$$

Se puede demostrar la existencia de una función de autocorrelación de este tipo, en series de tiempo que se obtienen a partir de la aplicación de un ajuste exponencial suave basado en la siguiente relación recursiva:

$$\begin{aligned} x_0 &= (1 - \lambda)r_0 \\ x_t &= \lambda x_{t-1} + (1 - \lambda)r_t, \end{aligned} \quad (4-166)$$

donde los números r_t son variables mutuamente independientes con media nula y variancia σ^2 .

Para propósitos prácticos, los valores r_t se pueden generar de manera tan simple como

se generan los números aleatorios uniformes transformados al intervalo $(-1/2, +1/2)$. En este caso, los valores de variable aleatoria x_i autocorrelacionados que han sido generados, tendrán un valor esperado cero y una variancia igual a

$$\sigma_x^2 \frac{1 - \lambda}{1 + \lambda} \sigma^2 = \frac{1 - \lambda}{12(1 + \lambda)}. \quad (4-167)$$

En la literatura sobre ajustes exponenciales suaves son bien conocidos los programas de computadora que proporcionan este tipo de autocorrelación

1. SUBROUTINE AUTOOCR (K, RX, X)
2. DIMENSION RX(12)
3. L = 12 - K
4. X = 0.0
5. DO 7 I = 1, L
6. RX(I) = RX(I + K)
7. X = X + RX(I)
8. L = L + 1
9. DO 11 I = L, 12
10. RX(I) = RND (R)
11. X = X + RX(I)
12. X = X - 6.0
13. RETURN
14. END

Figura 4-29. Subrutina FORTRAN para generar valores de variable autocorrelacionados.

y generalmente su estructura es muy simple, ya que en la ecuación (4-166) tan sólo se requiere un valor para representar la información pasada [5, p. 172].

REFERENCIAS Y BIBLIOGRAFÍA

1. Aitchison, J. y Brown, J. A. C. *The Lognormal Distribution*, Cambridge: Cambridge University Press, 1957.
2. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. Nueva York: John Wiley and Sons, 1958.
3. Box, G. E. P. y Muller, M. E. "A Note on the Generation of Normal Deviates", *Annals of Mathematical Statistics*, XXIX (1958), 610-611.
4. Brockmeyer, E. H., Halstrom, L. y Jensen, A. *The Life and Works of A. K. Erlang*. Copenhagen: Copenhagen Telephone Company, 1948.

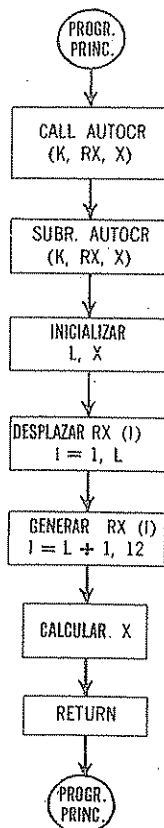


Figura 4-28. Diagrama de flujo para la generación de valores de variable aleatoria normales y autocorrelacionados.