

SIMULACIÓN

SHELDON M. ROSS

SEGUNDA EDICIÓN

En este texto se muestra la manera de analizar un modelo mediante un estudio de simulación. En particular, primero se explica la forma de utilizar una computadora para generar números aleatorios (más precisamente, pseudoaleatorios) y luego se ve la forma de emplearlos para generar los valores de variables aleatorias a partir de distribuciones arbitrarias.

Mediante el concepto de eventos discretos se muestra la manera de emplear variables aleatorias y generar el comportamiento de un modelo estocástico conforme transcurre el tiempo. Al hacerlo, estudiamos el modo de obtener estimadores de las cantidades que nos interesen. Se analizan las interrogantes estadísticas de cuándo detener una simulación y qué grado de confianza debe tenerse en los estimadores que resulten. Presentamos varias formas de mejorar los estimadores comunes de la simulación. Además, se plantea la forma de utilizarla para determinar si el modelo estocástico elegido es congruente con un conjunto de datos reales.

CONTENIDO

- Introducción
- Elementos de probabilidad
- Números aleatorios
- Generación de variables aleatorias discretas
- Generación de variables aleatorias continuas
- El método de simulación por medio de eventos discretos
- Análisis estadístico de datos simulados
- Técnicas de reducción de varianza
- Técnicas de validación estadística
- Métodos de Monte Carlo con cadenas de Markov
- Algunos temas adicionales

ROSS

SIMULACIÓN

SEGUNDA
EDICIÓN

PEARSON

PRENTICE
HALL

PEARSON

PRENTICE
HALL

SIMULACIÓN

SEGUNDA EDICIÓN

PEARSON

PRENTICE
HALL

SHELDON M. ROSS

70-17-0259-X

90000



702598

Visítenos en Internet:
<http://www.prentice.com.mx>
<http://www.prenhall.com>

Simulación

Segunda edición

Sheldon M. Ross

DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

TRADUCCIÓN:

ÓSCAR ALFREDO PALMAS VELASCO
Doctor en Matemáticas
Instituto de Matemáticas, UNAM

REVISIÓN TÉCNICA:

VÍCTOR HUGO IBARRA MERCADO
Lic. en Física y Matemáticas, ESFM-IPN
Catedrático de la Escuela de Actuaría,
Universidad Anáhuac

PEARSON

**PRENTICE
HALL**

**Addison
Wesley
Longman**

MÉXICO • ARGENTINA • BOLIVIA • BRASIL • COLOMBIA • COSTA RICA • CHILE • ECUADOR
EL SALVADOR • ESPAÑA • GUATEMALA • HONDURAS • NICARAGUA • PANAMÁ
PARAGUAY • PERÚ • PUERTO RICO • REPÚBLICA DOMINICANA • URUGUAY • VENEZUELA

AMSTERDAM • HARLOW • MIAMI • MUNICH • NUEVA DELHI • MENLO PARK • NUEVA JERSEY
NUEVA YORK • ONTARIO • PARÍS • SINGAPUR • SYDNEY • TOKIO • TORONTO • ZURICH

Ross, Sheldon M.
Simulación, 2a. edición

PRENTICE HALL, MÉXICO, 1999

ISBN: 970-17-0259-X

ÁREA: UNIVERSITARIOS

FORMATO: 17 x 23 cm

PÁGINAS: 296

EDICIÓN EN ESPAÑOL:

EDITOR: PABLO EDUARDO ROIG VÁZQUEZ
SUPERVISOR DE TRADUCCIÓN: JOSÉ LUIS NÚÑEZ HERREJÓN
SUPERVISOR DE EDICIÓN: ALEJANDRO A. GÓMEZ RUIZ
CORRECTOR DE ESTILO: JOSÉ FRANCISCO JAVIER DÁVILA MARTÍNEZ

ROSS: SIMULATION, 2a. ed.

Traducido de la segunda edición en inglés de la obra: SIMULATION.

All rights reserved. Authorized translation from English language edition published by Academic Press, Inc.

Todos los derechos reservados. Traducción autorizada de la edición en inglés publicada por Academic Press, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.

Prohibida la reproducción total o parcial de esta obra, por cualquier medio método sin autorización por escrito del editor.

Derechos reservados © 1999 respecto a la primera edición en español publicada por:

PRENTICE HALL HISPANOAMERICANA, S. A.
Calle 4 Núm. 25, 2º piso, Fracc. Industrial Alce Blanco
53370 Naucalpan de Juárez, Edo. de México

ISBN 970-17-0259-X

Miembro de la Cámara Nacional de la Industrial Editorial, Reg. Núm. 1524.

Original English Language Edition Published by
Academic Press, Inc.
Copyright © 1997
All rights reserved

ISBN 0-12-598410-3

IMPRESO EN MÉXICO/PRINTED IN MEXICO

Contenido

Prefacio

ix

Capítulo 1 Introducción

Ejercicios

3

Capítulo 2 Elementos de probabilidad

2.1 Espacio muestral y eventos	5
2.2 Axiomas de probabilidad	6
2.3 Probabilidad condicional e independencia	7
2.4 Variables aleatorias	8
2.5 Esperanza	10
2.6 Varianza	13
2.7 Desigualdad de Chebyshev y las leyes de los grandes números	15
2.8 Algunas variables aleatorias discretas	17
Variables aleatorias binomiales, 17 / Variables aleatorias Poisson, 19	
Variables aleatorias geométricas, 20 / La variable aleatoria binomial negativa, 21 / Variables aleatorias hipergeométricas, 21	
2.9 Variables aleatorias continuas	22
Variables aleatorias distribuidas uniformemente, 22 / Variables aleatorias normales, 23	
Variables aleatorias exponenciales, 25 / El proceso Poisson y las variables aleatorias gamma, 26 / El proceso Poisson no homogéneo, 29	
2.10 Esperanza condicional y varianza condicional	30
Ejercicios	31
Bibliografía	35

Capítulo 3 Números aleatorios

Introducción	36
3.1 Generación de números pseudoaleatorios	36
3.2 Uso de números aleatorios para evaluar integrales	38
Ejercicios	42
Bibliografía	44

Capítulo 4 Generación de variables aleatorias discretas

4.1 El método de la transformada inversa	45
4.2 Generación de una variable aleatoria Poisson	50
4.3 Generación de variables aleatorias binomiales	52
4.4 La técnica de aceptación y rechazo	53
4.5 El método de composición	56
Ejercicios	57

Capítulo 5 Generación de variables aleatorias continuas

Introducción	62
5.1 El algoritmo de la transformada inversa	62
5.2 El método de rechazo	66
5.3 El método polar para generar variables aleatorias normales	72
5.4 Generación de un proceso Poisson	75
5.5 Generación de un proceso Poisson no homogéneo	77
Ejercicios	81
Bibliografía	85

Capítulo 6 El método de simulación por medio de eventos discretos

Introducción	86
6.1 Simulación mediante eventos discretos	86
6.2 Sistema de línea de espera con un servidor	88
6.3 Sistema de línea de espera con dos servidores en serie	90
6.4 Sistema de línea de espera con dos servidores en paralelo	94
6.5 Modelo de inventario	97
6.6 Problema de reparación	99
6.7 Ejercicio de opciones en acciones	101
6.8 Verificación del modelo de simulación	104
Ejercicios	106
Bibliografía	109

LA BIBLIOTECA
DE LA UNIVERSIDAD
DE LOS ANGELES

Capítulo 7 Análisis estadístico de datos simulados

Introducción	111
7.1 Media y varianza muestrales	112
7.2 Estimación del intervalo de una media poblacional	117
7.3 La técnica bootstrap para estimar errores cuadráticos medios	121
Ejercicios	128
Bibliografía	130

Capítulo 8 Técnicas de reducción de varianza

Introducción	131
8.1 El uso de variables antitéticas	133
8.2 El uso de variables de control	141
8.3 Reducción de varianza mediante condicionamiento	149
8.4 Muestreo estratificado	158
8.5 Muestreo de importancia	167
8.6 Uso de números aleatorios comunes	178
Apéndice: verificación del método de las variables antitéticas al estimar el valor esperado de funciones monótonas	180
Ejercicios	182
Bibliografía	188

Capítulo 9 Técnicas de validación estadística

Introducción	189
9.1 Pruebas de bondad de ajuste	189
Prueba de bondad de ajuste ji-cuadrada para datos discretos, 190 / Prueba de Kolmogorov-Smirnov para datos continuos, 193	
9.2 Pruebas de bondad de ajuste sin parámetros especificados	198
El caso de los datos discretos, 198 / El caso de los datos continuos, 201	
9.3 El problema de las dos muestras	202
9.4 Validación de la hipótesis de un proceso Poisson no homogéneo	210
Ejercicios	214
Bibliografía	216

Capítulo 10 Métodos de Monte Carlo con cadenas de Markov

Introducción	218
10.1 Cadenas de Markov	218
10.2 El algoritmo de Hastings-Metropolis	221
10.3 El muestreador de Gibbs	224
10.4 Temple simulado	236
10.5 El algoritmo de muestreo con remuestreo de importancia	239
Ejercicios	243
Bibliografía	

ESTE DOCUMENTO
PERTENECE A:
MANUEL ALONSO CASTRO GIL

Capítulo 11 Algunos temas adicionales

Introducción	248
11.1 El método de las variables alias para generar variables aleatorias discretas	248
11.2 Simulación de un proceso Poisson bidimensional	252
11.3 Aplicaciones de una identidad para la suma de variables aleatorias Bernoulli a la simulación	256
11.4 Estimación de probabilidades y tiempos esperados de primera pasada mediante riesgos aleatorios	260
Ejercicios	267
Bibliografía	268
<i>Apéndice de programas</i>	270
<i>Índice</i>	279



Prefacio

Al formular un modelo estocástico para describir un fenómeno real, lo común era hallar el término medio entre elegir un modelo que fuese una imagen realista de la situación y uno cuyo análisis matemático resultara posible. No parecía adecuado elegir, pues, un modelo fiel al fenómeno en estudio si no era posible analizarlo desde el punto de vista matemático. Ciertas consideraciones similares han llevado a concentrarse en resultados asintóticos o de estado estacionario, en contraposición a los más útiles relativos al tiempo transitorio. Sin embargo, la aparición más o menos reciente de veloces y no muy caras computadoras ha llevado a otro planteamiento, a saber: tratar de modelar el fenómeno de la manera más fiel posible y luego basarse en un estudio de simulación para analizarlo.

En este texto mostramos la forma de analizar un modelo mediante un estudio de simulación. En particular, primero explicamos la forma de utilizar una computadora para generar números aleatorios (más precisamente, pseudoaleatorios) y luego ver la forma de utilizarlos para generar los valores de variables aleatorias a partir de distribuciones arbitrarias. Nos servimos del concepto de eventos discretos para mostrar la forma de emplear variables aleatorias y generar el comportamiento de un modelo estocástico conforme transcurre el tiempo. Al hacerlo, estudiamos el modo de obtener estimadores de las cantidades que nos interesen. Analizamos las interrogantes estadísticas de cuándo detener una simulación y qué grado de confianza debe tenerse en los estimadores que resulten. Presentamos varias formas de mejorar los estimadores comunes de la simulación. Además, mostramos la forma de utilizarla para determinar si el modelo estocástico elegido es congruente con un conjunto de datos reales.

Los capítulos del libro son los siguientes: el capítulo 1 es introductorio y presenta un fenómeno característico de interés. El capítulo 2 es un repaso de probabilidad. Aunque es completo y no supone que el lector está familiarizado con esa materia, imaginamos que de hecho será un repaso para muchos. El capítulo 3 trata de los

números aleatorios y de cómo una variante de ellos (los llamados números pseudo-aleatorios) pueden generarse en una computadora. En los capítulos 4 y 5 se considera el uso de números aleatorios para generar variables aleatorias discretas y luego continuas.

El capítulo 6 presenta el método de eventos discretos para el seguimiento de un sistema arbitrario conforme éste evoluciona con el tiempo. Se presentan varios ejemplos, relacionados con los sistemas de líneas de espera con un solo servidor y con varios servidores, con un sistema de inventarios, con un modelo de reparación de una máquina y para ejercer una opción de acciones. El capítulo 7 presenta el tema de la estadística. Suponemos que nuestro lector promedio no ha estudiado el tema, por lo cual partimos de conceptos muy básicos y al final presentamos el método estadístico de "bootstrap", que es muy útil para analizar los resultados de una simulación.

El capítulo 8 estudia el importante tema de reducción de la varianza. Se trata de una tentativa por mejorar los estimadores comunes de una simulación, determinando los que tengan la misma media pero menor varianza. El capítulo comienza con la técnica de uso de variables antitéticas. Observamos (transfiriendo la demostración a un apéndice) que esto siempre produce una reducción de la varianza, junto con un ahorro computacional, cuando intentamos estimar el valor esperado de una función que es monótona en cada una de sus variables. Luego presentamos las variables de control e ilustramos su utilidad en la reducción de la varianza. Por ejemplo, mostramos la forma de utilizar las variables de control de manera eficaz para analizar los sistemas de líneas de espera, los sistemas de fiabilidad, un problema de reordenamiento de listas y el *blackjack*. También indicamos cómo emplear paquetes de regresión para facilitar los cálculos que resultan de aprovechar las variables de control. Luego se estudia la reducción de varianza mediante el uso de esperanzas condicionales. Tal uso se indica en ejemplos que tratan la estimación de π y el análisis de los sistemas de líneas de espera con capacidad finita. Además, junto con una variable de control, la esperanza condicional sirve para estimar el número esperado de eventos de un proceso de renovación durante cierto tiempo fijo. El muestreo estratificado como una herramienta para la reducción de varianza se indica en ejemplos de líneas de espera con tasas de llegada variables y de evaluación de integrales. La relación entre las técnicas de reducción de varianza de esperanza condicional y muestreo estratificado se explica e ilustra en la estimación de la ganancia esperada en el video póquer. Después analizamos la técnica de muestreo de importancia. Indicamos y explicamos cómo ésta puede ser una técnica muy poderosa en la reducción de varianza al estimar probabilidades pequeñas. Al hacerlo, presentamos el concepto de distribuciones inclinadas y mostramos la forma de utilizarlas en una estimación con muestreo de importancia de una pequeña probabilidad con cola de convolución. Se presentan aplicaciones del muestreo de importancia a las líneas de espera, las caminatas aleatorias, las permutaciones aleatorias y al cálculo de esperanzas condicionales cuando una está condicionando un evento raro. La última

técnica de reducción de varianza del capítulo 8 se relaciona con el uso de un flujo común de números aleatorios.

El capítulo 9 trata de las técnicas de validación estadística, que son procedimientos estadísticos útiles para validar el modelo estocástico cuando se dispone de algunos datos reales. Se presentan pruebas de bondad de ajuste, como la prueba ji cuadrada y la prueba de Kolmogorov-Smirnov. Otras secciones del capítulo tratan los problemas de dos muestras y n muestras y formas de verificar estadísticamente la hipótesis de que un proceso sea Poisson.

El capítulo 10 trata de los métodos de Monte Carlo con cadenas de Markov. Se trata de técnicas que han ampliado en gran medida el uso de la simulación en los últimos años. El paradigma canónico de simulación para estimar $\theta = E[h(\mathbf{X})]$, donde \mathbf{X} es un vector aleatorio, consiste en simular copias independientes e idénticamente distribuidas de \mathbf{X} y luego servirse del valor promedio de $h(\mathbf{X})$ como estimador. Éste es el estimador de simulación "bruto", que luego se puede mejorar mediante una o más de las ideas de reducción de varianza del capítulo 8. Sin embargo, con este método es necesario especificar la distribución de \mathbf{X} y que se pueda realizar una simulación. No obstante, como veremos en el capítulo 10, hay muchos ejemplos en los que se conoce la distribución de \mathbf{X} pero no se puede simular en forma directa el vector aleatorio \mathbf{X} , así como otros ejemplos en los que la distribución no se conoce de manera completa, sino que se especifica salvo una constante multiplicativa. Así, en estos casos no se dispone del método usual para estimar θ . Sin embargo, en los últimos años se ha utilizado un nuevo método, basado en la generación de una cadena de Markov cuya distribución límite es la distribución de \mathbf{X} , a lo que sigue la estimación de θ mediante el promedio de los valores de la función h evaluada en los estados sucesivos de la cadena. Estos métodos de Monte Carlo con cadenas de Markov se exploran en el capítulo 10. Comenzamos, en la sección 10.2, presentando algunas propiedades de las cadenas de Markov. En la sección 10.3 se presenta una técnica general para generar una cadena de Markov con una distribución límite dada salvo una constante multiplicativa, conocida como algoritmo de Hastings-Metropolis, y se da una aplicación para generar un elemento aleatorio de un enorme conjunto "combinatorio". La versión de uso más amplio de este algoritmo es el muestreador de Gibbs, que se presenta en la sección 10.4. Se analizan ejemplos relativos a la generación de puntos aleatorios en una región, sujetos a la restricción de que nunca dos puntos están a menos de una distancia fija uno de otro; para el análisis de redes de líneas de espera de productos, y para el análisis de un modelo jerárquico de estadística bayesiana para predecir el número de cuadrangulares que lograrán ciertos jugadores de béisbol. En la sección 10.5 se presenta una aplicación de los métodos de este capítulo a problemas de optimización determinista, llamados de temple simulado, así como un ejemplo relativo al problema del agente de ventas viajero. La última sección del capítulo 10 trata del algoritmo de muestreo con remuestreo de importancia, que es una gene-

realización de la técnica de aceptación y rechazo de los capítulos 4 y 5. Se explica el uso de este algoritmo en estadística bayesiana.

El capítulo 11 toca algunos otros temas de la simulación. En la sección 11.1 revisamos el método de variables alias, el cual, con cierto costo de configuración, es una forma muy eficaz de generar variables aleatorias discretas. La sección 11.2 trata la simulación de un proceso Poisson bidimensional. En la sección 11.3 presentamos una identidad relativa a la covarianza de la suma de variables aleatorias Bernoulli dependientes y mostramos que su uso puede producir estimadores de pequeñas probabilidades con muy pequeñas varianzas. Esto se aplica para estimar la fiabilidad de los sistemas, lo que parece ser más eficaz que cualquier otro estimador conocido de la fiabilidad de un sistema pequeño, así como para estimar la probabilidad de que determinado patrón aparezca durante cierto tiempo fijo. La sección 11.4 presenta los riesgos aleatorios. Estas cantidades se pueden emplear como variables de control al estimar el tiempo medio transcurrido hasta que un determinado proceso de Markov alcance un conjunto específico de estados. Además, en situaciones en las que no sea seguro que se alcance el conjunto de estados, el riesgo aleatorio sirve como estimador de la probabilidad de este evento; es un estimador de particular eficiencia cuando la probabilidad de llegar al conjunto de estados sea pequeña.

CAPÍTULO

1

Introducción

Consideremos la situación que enfrenta cierto farmacéutico que piensa en instalar una pequeña farmacia para surtir recetas. Planea abrir todos los días a las nueve de la mañana y espera que, en promedio, surtirá alrededor de 32 recetas antes de las cinco de la tarde. Su experiencia le ha mostrado que el tiempo que tarda en surtir una receta, una vez recibida, es una cantidad aleatoria con una media y desviación estándar de 10 y cuatro minutos, respectivamente. Planea no aceptar más recetas después de las cinco, aunque se quedará en la tienda más tiempo si hiciera falta para cumplir los pedidos del día. Es probable que dado este panorama, el farmacéutico esté interesado en responder las siguientes preguntas:

1. En promedio, ¿a qué hora de la tarde saldrá de su negocio?
2. ¿Cuál será la proporción de días en que seguirá trabajando a las cinco y media de la tarde?
3. ¿Cuál es el tiempo promedio que tardará en surtir una receta (tomando en cuenta que no necesariamente comenzará a trabajar de inmediato con un nuevo pedido, sino que tendrá que concluir antes con las anteriores)?
4. ¿Qué proporción de recetas surtirá en un lapso de 30 minutos?
5. Si modifica su política de aceptar todas las recetas entre las nueve y las cinco, y en cambio sólo recibe nuevos pedidos cuando tenga menos de cinco recetas pendientes, ¿cuántas recetas perderá en promedio?
6. ¿Cómo influiría la restricción anterior en las respuestas de las preguntas 1 a 4?

Con el fin de utilizar las matemáticas para analizar esta situación y dar respuesta a las preguntas, primero construimos un modelo de probabilidad. Para ello, es razonable plantear varias hipótesis relativas a nuestro marco de referencia. Por ejemplo,

debemos hacer ciertas suposiciones relativas al mecanismo probabilístico que describe las llegadas del promedio diario de 32 clientes. Una hipótesis posible es que la tasa de llegada sea, en un sentido probabilístico, constante durante el día, mientras que otra (tal vez más realista) es que esta tasa dependa de la hora. Entonces, debemos especificar una distribución de probabilidad (con media 10 y desviación estándar 4) para el tiempo que toma surtir una receta y establecer hipótesis relativas al hecho de que el tiempo de servicio de una receta dada siempre tenga esta distribución o dependa de otras variables (por ejemplo, el número de recetas pendientes por surtir o la hora del día). Es decir, debemos plantear hipótesis probabilísticas en torno de los tiempos de llegada y de servicio diarios. También tenemos que decidir si la ley de probabilidad que describe los cambios en un día determinado dependen del día de la semana, o bien si básicamente permanecen constantes en el tiempo. Después de especificar estas hipótesis y tal vez otras más, se habrá formulado un modelo de probabilidad para esta situación.

Ya con un modelo de probabilidad, es posible, en teoría, responder de manera analítica a las preguntas. Sin embargo, en la práctica estas cuestiones son demasiado difíciles para ser determinadas de esta manera; así, para contestarlas por lo general debemos realizar un estudio de simulación, es decir, uno que programa el mecanismo probabilístico en una computadora y se vale de “números aleatorios” para simular las posibles ocurrencias del modelo durante muchos días, para luego aplicar la teoría de la estadística y estimar las respuestas a las preguntas planteadas. En otras palabras, el programa de computadora utiliza números aleatorios para generar los valores de variables aleatorias con la distribución de probabilidad supuesta, que representen los tiempos de llegada y de servicio de las recetas. Con estos valores, para muchos días, determina las cantidades de interés relacionadas con estas cuestiones. Luego recurre a las técnicas estadísticas para ofrecer estimaciones; por ejemplo, si de 1000 días simulados en 122 el farmacéutico continúa trabajando a las cinco y media, estimaríamos que la respuesta a la pregunta 2 es 0.122.

Para realizar este análisis, uno debe tener ciertos conocimientos de probabilidad a fin de decidir sobre cierta distribución de probabilidad y acerca de cuestiones como suponer o no la independencia de las variables aleatorias adecuadas. Por eso, en el capítulo 2 aparece un repaso de probabilidad. Las bases de un estudio de simulación son los llamados números aleatorios. En el capítulo 3 se presenta un análisis de estas cantidades y la forma de generarlas por computadora. Los capítulos 4 y 5 muestran la forma de utilizar los números aleatorios para generar los valores de variables aleatorias con distribuciones arbitrarias. Las distribuciones discretas se estudian en el capítulo 4 y las continuas en el capítulo 5. Al concluir ese capítulo, el lector habrá adquirido cierta intuición acerca de la formulación de un modelo de probabilidad para un sistema dado y también sabrá cómo emplear los números aleatorios para generar los valores de cantidades aleatorias relacionadas con tal modelo. En el capítulo 6 se aprove-

chan estos valores generados para realizar un seguimiento del sistema con el paso del tiempo (es decir, la simulación del sistema), y ahí mismo presentamos el concepto de “eventos discretos” e indicamos la forma de servirnos de ellos a fin de lograr un método sistemático para simular sistemas. El método de simulación por medio de eventos discretos conduce a un programa de computadora, el cual se puede escribir en el lenguaje que resulte más cómodo para el lector, y que simule al sistema un gran número de veces. En el capítulo 6 también daremos algunas sugerencias en torno de la verificación del programa, para asegurarnos de que realmente hace lo que debe hacer. El uso de las salidas de un estudio de simulación para responder preguntas probabilísticas relativas al modelo necesita de la teoría de la estadística, tema que se presenta en el capítulo 7. Este capítulo comienza con los conceptos más elementales y sencillos de la estadística y avanza hacia la reciente innovación de la técnica de “bootstrap”, que es muy útil en la simulación. Nuestro estudio de la estadística indica la importancia de la varianza de los estimadores obtenida a partir del estudio de simulación como indicador de la eficiencia de ésta. En particular, mientras más pequeña sea la varianza, menor será la cantidad de simulación necesaria para obtener una precisión fija. Como resultado, en el capítulo 8 veremos formas de producir nuevos estimadores que sean mejores que los estimadores de simulación en bruto, pues tendrán menores varianzas. El tema de reducción de la varianza es muy importante en un estudio de simulación, pues puede mejorar su eficiencia. El capítulo 9 muestra cómo aplicar los resultados de la simulación para verificar qué tan adecuado es el modelo de probabilidad (simulado) en la situación del mundo real, cuando se dispone de datos reales. El capítulo 10 presenta el importante tema de los métodos de Monte Carlo con cadenas de Markov. En los últimos años, estos métodos han ampliado la gama de problemas que se enfrentan mediante simulación. El capítulo 11 se dedica a varios temas adicionales.

Ejercicios

1. Los siguientes datos proporcionan los tiempos de llegada y de servicio que se invierten en cada cliente, para los primeros 13 clientes en un sistema con un único servidor.

Al llegar un cliente, pasa a servicio si el servidor está libre o bien se forma en la línea de espera. Cuando el servidor concluye el trabajo de un cliente, el siguiente de la línea (es decir, quien ha esperado más tiempo) entra a servicio.

Tiempos de llegada: 12 31 63 95 99 154 198 221 304 346 411 455 537
 Tiempos de servicio: 40 32 55 48 18 50 47 18 28 54 40 72 12

- (a) Determine los tiempos de salida de estos primeros 13 clientes.
 - (b) Repita (a) considerando que hay dos servidores que atienden a los clientes.
 - (c) Repita (a) con la nueva hipótesis de que cuando el servidor concluye un servicio, el siguiente cliente debe ser el que haya esperado menos tiempo.
2. Considere una estación de servicio en la que los clientes son atendidos por orden de llegada. Sean A_n , S_n y D_n , respectivamente, la hora de llegada, el tiempo de servicio y la hora de salida del cliente n . Suponga que sólo hay un servidor y que el sistema inicialmente no tiene clientes.
- (a) Argumente que $D_0 = 0$, y para $n > 0$,

$$D_n - S_n = \text{Máximo} \{A_n, D_{n-1}\}.$$

- (b) Determine la fórmula de recursión correspondiente cuando hay dos servidores.
- (c) Determine la fórmula de recursión correspondiente cuando hay k servidores.
- (d) Escriba un programa de computadora que determine las horas de salida en función de la hora de llegada y del tiempo de servicio. Utilice el programa para verificar sus respuestas de la parte (a) y (b) del ejercicio 1.

CAPÍTULO

2

Elementos
de probabilidad

2.1 Espacio muestral y eventos

Consideremos un experimento cuyo resultado no se conoce de antemano. Sea S el espacio muestral del experimento, es decir, el conjunto de todos los resultados posibles. Por ejemplo, si el experimento es una carrera de caballos numerados del 1 al 7, entonces

$$S = \{\text{todos los ordenamientos de } (1, 2, 3, 4, 5, 6, 7)\}$$

Por ejemplo, el resultado $(3, 4, 1, 7, 6, 5, 2)$ significa que el caballo número 3 llegó en primer lugar, el caballo número 4 llegó en segundo, y así sucesivamente.

Cualquier subconjunto A del espacio muestral se conoce como evento, lo que significa que un evento es un conjunto formado por resultados posibles del experimento. Si el resultado del experimento está en A , decimos que A ha ocurrido. En el ejemplo anterior, si

$$A = \{\text{todos los resultados de } S \text{ que comienzan con } 5\}$$

entonces A es el evento de que el caballo número 5 llegue en primer lugar.

Para cualesquiera dos eventos A y B , definimos el nuevo evento $A \cup B$, llamado unión de A y B , consistente en todos los resultados que están en A o en B o en ambos. De manera similar, definimos el evento AB , llamado intersección de A y B , como todos los resultados que están en A y en B . Es decir, el evento $A \cup B$ ocurre si A o B ocurre, mientras que el evento AB ocurre si ambos ocurren. También podemos definir uniones e intersecciones de más de dos eventos. En particular, la unión de los eventos A_1, A_2, \dots, A_n (denotada por $\cup_{i=1}^n A_i$) se define como la formada por todos los resultados que están en cualquiera de los A_i . De manera análoga, la intersección de los

eventos A_1, A_2, \dots, A_n (denotada $A_1 A_2 \cdots A_n$) se define como la formada por todos los resultados que están en todos los A_i .

Para cualquier evento A , definimos el evento A^c , el complemento de A , como el que consiste en todos los resultados del espacio muestral S que no están en A . Es decir, A^c ocurre si y sólo si A no ocurre. Como el resultado de un experimento debe estar en el espacio muestral S , esto implica que S^c no contiene resultado alguno y por lo tanto no puede ocurrir. Llamamos a S^c el conjunto vacío y lo denotamos por \emptyset . Si $AB = \emptyset$, de modo que A y B no pueden ocurrir a la vez (pues no hay resultados que estén en A y en B), decimos que A y B son mutuamente excluyentes.

2.2 Axiomas de probabilidad

Supongamos que para cada evento A de un experimento con espacio muestral S hay un número, denotado $P(A)$ y llamado la probabilidad del evento A , que cumple los tres axiomas siguientes:

Axioma 1 $0 \leq P(A) \leq 1$

Axioma 2 $P(S) = 1$

Axioma 3 Para cualquier secuencia de eventos mutuamente excluyentes A_1, A_2, \dots

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \quad n = 1, 2, \dots, \infty$$

Así, el axioma 1 establece que la probabilidad de que el resultado del experimento esté en A es cierto número entre 0 y 1. El axioma 2 establece que con probabilidad 1, este resultado es un elemento del espacio muestral; y el axioma 3 afirma que para cualquier conjunto de eventos mutuamente excluyentes, la probabilidad de que al menos uno ocurra es igual a la suma de sus probabilidades respectivas.

Estos tres axiomas pueden utilizarse para demostrar varios resultados relativos a las probabilidades. Por ejemplo, como A y A^c siempre son eventos mutuamente excluyentes, y como $A \cup A^c = S$, tenemos de los axiomas 2 y 3 que

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$$

o en forma equivalente,

$$P(A^c) = 1 - P(A)$$

En palabras, la probabilidad de que no ocurra un evento es 1 menos la probabilidad de que ocurra.

2.3 Probabilidad condicional e independencia

Consideremos un experimento consistente en lanzar al aire una moneda dos veces y observar si el resultado es cara o cruz. El espacio muestral puede considerarse como el siguiente conjunto de cuatro resultados:

$$S = \{(cara, cara), (cara, cruz), (cruz, cara), (cruz, cruz)\}$$

donde (cara, cruz) significa, por ejemplo, que en el primer lanzamiento se obtiene cara y en el segundo cruz. Supongamos ahora que cada uno de los cuatro posibles resultados es igualmente probable y que por lo tanto tiene $\frac{1}{4}$ de probabilidad. Supongamos además que observamos que el primer lanzamiento cae en cara. Entonces, dada esta información, ¿cuál es la probabilidad de que ambos lanzamientos caigan en cara? Para calcular esta probabilidad, razonamos de la manera siguiente: dado que el lanzamiento inicial cayó en cara, pueden existir a lo más dos resultados posibles, a saber, (cara, cara) o (cara, cruz). Además, como cada uno de estos resultados tenía originalmente la misma probabilidad de ocurrir, deberían seguir teniendo probabilidades iguales. Es decir, dado que el primer lanzamiento cayó en cara, la probabilidad (condicional) de cada uno de los resultados (cara, cara) y (cara, cruz) es de $\frac{1}{2}$, mientras que la probabilidad (condicional) de los otros dos resultados es de 0. Por lo tanto, la probabilidad deseada es $\frac{1}{2}$.

Si A y B denotan, respectivamente, los eventos de que ambos lanzamientos caigan en cara y el evento de que el primer lanzamiento caiga en cara, entonces la probabilidad obtenida anteriormente es la probabilidad condicional de A dado que B ha ocurrido y se denota como

$$P(A|B)$$

Podemos obtener una fórmula general para $P(A|B)$, válida para todos los experimentos y eventos A y B , como antes: si el evento B ocurre, entonces para que A ocurra es necesario que lo ocurrido sea un punto en A y en B ; es decir, debe estar en AB . Ahora, como sabemos que B ha ocurrido, esto implica que B se convierte en nuestro nuevo espacio muestral y por lo tanto la probabilidad de que el evento AB ocurra será igual a la probabilidad de AB con respecto de la probabilidad de B . Es decir,

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Como indica el ejemplo del lanzamiento de monedas, $P(A|B)$, la probabilidad condicional de A , dado que B ha ocurrido, no es en general igual a $P(A)$, la probabilidad no condicional de A . En otras palabras, el hecho de saber que B ha ocurrido

modifica por lo general la probabilidad de que A ocurra (¿qué pasaría si fuesen mutuamente excluyentes?). En el caso particular en que $P(A|B)$ sea igual a $P(A)$, decimos que A y B son independientes. Como $P(A|B) = P(AB)/P(B)$, vemos que A es independiente de B si

$$P(AB) = P(A)P(B)$$

Como esta relación es simétrica en A y B , esto implica que siempre que A sea independiente de B , B es independiente de A .

2.4 Variables aleatorias

Cuando se realiza un experimento, a veces nos interesa principalmente cierta cantidad numérica determinada por el resultado. Estas cantidades de interés que son determinadas por los resultados del experimento se conocen como variables aleatorias.

La función de distribución acumulativa, o más brevemente, la función de distribución F de la variable aleatoria X se define para cualquier número real x como

$$F(x) = P\{X \leq x\}$$

Una variable aleatoria que puede asumir un número finito o a lo más una cantidad numerable de valores posibles es discreta. Para una variable aleatoria discreta X , definimos su función de masa de probabilidad $p(x)$ como

$$p(x) = P\{X = x\}$$

Si X es una variable aleatoria discreta que asume uno de los posibles valores x_1, x_2, \dots , entonces, como X debe tomar uno de estos valores, tenemos

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Ejemplo 2a Supongamos que X toma uno de los valores 1, 2 o 3. Si

$$p(1) = \frac{1}{4}, \quad p(2) = \frac{1}{3}$$

entonces, como $p(1) + p(2) + p(3) = 1$, esto implica que $p(3) = \frac{5}{12}$. ■

Mientras una variable aleatoria discreta asume a lo más un conjunto numerable de posibles valores, con frecuencia debemos analizar variables aleatorias cuyo conjunto de posibles valores es un intervalo. Decimos que la variable aleatoria X es una varia-

ble aleatoria continua si existe una función no negativa $f(x)$ definida para todo número real x , con la propiedad de que para cualquier conjunto C de números reales

$$P\{X \in C\} = \int_C f(x) dx \quad (2.1)$$

La función f es la función de densidad de probabilidad de la variable aleatoria X .

La relación entre la distribución acumulativa $F(\cdot)$ y la densidad de probabilidad $f(\cdot)$ se expresa como

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) dx$$

Al derivar ambos lados se obtiene

$$\frac{d}{da} F(a) = f(a)$$

Es decir, la densidad es la derivada de la función de distribución acumulativa. Una interpretación un poco más intuitiva de la función de densidad se obtiene de la ecuación (2.1) como sigue:

$$P\left\{a - \frac{\epsilon}{2} \leq X \leq a + \frac{\epsilon}{2}\right\} = \int_{a-\epsilon/2}^{a+\epsilon/2} f(x) dx \approx \epsilon f(a)$$

donde ϵ es pequeño. En otras palabras, la probabilidad de que X esté contenido en un intervalo de longitud ϵ alrededor del punto a es aproximadamente $\epsilon f(a)$. Esto permite ver que $f(a)$ es una medida de qué tan probable es que la variable aleatoria esté cerca de a .

En muchos experimentos estamos interesados no sólo en las funciones de distribución de probabilidad de variables aleatorias individuales, sino también en las relaciones entre dos o más de ellas. Para especificar la relación entre dos variables aleatorias, definimos la función de distribución de probabilidad acumulativa conjunta de X y Y como

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

Así, $F(x, y)$ especifica la probabilidad de que X sea menor o igual a x y que en forma simultánea Y sea menor o igual a y .

Si tanto X como Y son variables aleatorias discretas, entonces definimos la función de masa de probabilidad conjunta de X y Y como

$$p(x, y) = P\{X = x, Y = y\}$$

De manera análoga, decimos que X y Y son conjuntamente continuas, con función de densidad de probabilidad conjunta $f(x, y)$, si para cualesquiera conjuntos de números reales C y D

$$P\{X \in C, Y \in D\} = \iint_{\substack{x \in C \\ y \in D}} f(x, y) dx dy$$

Las variables aleatorias X y Y son independientes si para cualesquiera dos conjuntos de números reales C y D

$$P\{X \in C, Y \in D\} = P\{X \in C\}P\{Y \in D\}$$

Es decir, X y Y son independientes si para todos los conjuntos C y D los eventos $A = \{X \in C\}$, $B = \{Y \in D\}$ son independientes. Intuitivamente, X y Y son independientes si el hecho de conocer el valor de una de ellas no influye en la distribución de probabilidad de la otra. Las variables aleatorias que no son independientes son dependientes.

Los axiomas de probabilidad permiten mostrar que las variables aleatorias discretas X y Y son independientes si y sólo si, para toda x, y ,

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}$$

De manera análoga, si X y Y son conjuntamente continuas con función de densidad $f(x, y)$, entonces son independientes si y sólo si, para toda x, y ,

$$f(x, y) = f_X(x)f_Y(y)$$

donde $f_X(x)$ y $f_Y(y)$ son las funciones de densidad de X y Y , respectivamente.

2.5 Esperanza

Uno de los conceptos más útiles en probabilidad es el de la esperanza de una variable aleatoria. Si X es una variable aleatoria discreta que toma uno de los posibles valores x_1, x_2, \dots , entonces la *esperanza* o *valor esperado* de X , también llamado media de X y denotado por $E[X]$, se define como

$$E[X] = \sum_i x_i P\{X = x_i\} \quad (2.2)$$

En palabras, el valor esperado de X es un promedio ponderado de los valores que puede tomar X , donde el peso está dado por la probabilidad de que X lo tome. Por ejemplo, si la función de masa de probabilidad de X está dada por

$$p(0) = \frac{1}{2} = p(1)$$

entonces

$$E[X] = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

es simplemente el promedio común de los dos valores posibles 0 y 1, que puede tomar X . Por otro lado, si

$$p(0) = \frac{1}{3}, \quad p(1) = \frac{2}{3}$$

entonces

$$E[X] = 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right) = \frac{2}{3}$$

es un promedio ponderado de los dos valores posibles 0 y 1, donde el valor 1 tiene el doble del peso del valor 0, pues $p(1) = 2p(0)$.

Ejemplo 2b Si I es una variable aleatoria indicadora del evento A , es decir, si

$$I = \begin{cases} 1 & \text{si } A \text{ ocurre} \\ 0 & \text{si } A \text{ no ocurre} \end{cases}$$

entonces

$$E[I] = 1P(A) + 0P(A^c) = P(A)$$

Por lo tanto, la esperanza de la variable aleatoria indicadora para el evento A es simplemente la probabilidad de que A ocurra. ■

Si X es una variable aleatoria continua con función de densidad de probabilidad f , entonces, en analogía con la ecuación (2.2), definimos el valor esperado de X como

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Ejemplo 2c Si la función de densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 3x^2 & \text{si } 0 < x < 1 \\ 0 & \text{en caso contrario} \end{cases}$$

entonces

$$E[X] = \int_0^1 3x^3 dx = \frac{3}{4}$$

Supongamos ahora que queremos determinar el valor esperado, no de la variable aleatoria X , sino de la variable aleatoria $g(X)$, donde g es cierta función dada. Como $g(X)$ toma el valor $g(x)$ cuando X toma el valor x , parece claro intuitivamente que $E[g(X)]$ debe ser un promedio ponderado de los valores posibles $g(x)$, donde, para una x dada, el peso dado a $g(x)$ es igual a la probabilidad (o densidad de probabilidad en el caso continuo) de que X sea igual a x . De hecho, lo anterior se puede demostrar, por lo cual tenemos el siguiente resultado.

Proposición Si X es una variable aleatoria discreta con función de masa de probabilidad $p(x)$, entonces

$$E[g(X)] = \sum_x g(x)p(x)$$

mientras que, si X es continua con función de densidad de probabilidad $f(x)$, entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

Una consecuencia de la proposición anterior es el siguiente.

Corolario Si a y b son constantes, entonces

$$E[aX + b] = aE[X] + b$$

Demostración En el caso discreto

$$\begin{aligned} E[aX + b] &= \sum_x (ax + b)p(x) \\ &= a \sum_x xp(x) + b \sum_x p(x) \\ &= aE[X] + b \end{aligned}$$

Como la demostración en el caso continuo es similar, hemos demostrado el resultado. ■

Se puede mostrar que la esperanza es una operación lineal, en el sentido de que para cualesquiera variables aleatorias X_1 y X_2

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

lo cual se puede generalizar con facilidad para obtener

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i]$$

2.6 Varianza

Aunque $E[X]$, el valor esperado de la variable aleatoria, X , es un promedio ponderado de los valores posibles de X , no proporciona información alguna acerca de la variación de tales valores. Una forma de medir esta variación es considerar el valor promedio del cuadrado de la diferencia entre X y $E[X]$. Así, esto nos conduce a la siguiente definición.

Definición Si X es una variable aleatoria con media μ , entonces la varianza de X , denotada por $\text{Var}(X)$, se define como

$$\text{Var}(X) = E[(X - \mu)^2]$$

Obtenemos una fórmula alternativa para $\text{Var}(X)$ como sigue:

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

Es decir,

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Una identidad útil, cuya demostración se deja como ejercicio, es que para cualesquiera constantes a y b

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

En tanto que el valor esperado de una suma de variables aleatorias es igual a la suma de las esperanzas, en general, el resultado correspondiente no es válido para las

varianzas. Sin embargo, es cierto en el importante caso particular en que las variables aleatorias son independientes. Antes de demostrar esto, definimos el concepto de covarianza entre dos variables aleatorias.

Definición La covarianza de dos variables aleatorias X y Y , denotada $\text{Cov}(X, Y)$, se define como

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

donde $\mu_x = E[X]$ y $\mu_y = E[Y]$.

Para conseguir una útil expresión de $\text{Cov}(X, Y)$ desarrollamos el lado derecho de la ecuación anterior, haciendo uso de la linealidad de la esperanza. Esto implica que

$$\begin{aligned}\text{Cov}(X, Y) &= E[XY - \mu_x Y - X\mu_y + \mu_x \mu_y] \\ &= E[XY] - \mu_x E[Y] - E[X]\mu_y + \mu_x \mu_y \\ &= E[XY] - E[X]E[Y]\end{aligned}\quad (2.3)$$

Ahora deducimos una expresión para $\text{Var}(X + Y)$ en términos de sus varianzas individuales y la covarianza entre ellas. Como

$$E[X + Y] = E[X] + E[Y] = \mu_x + \mu_y$$

vemos que

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y - \mu_x - \mu_y)^2] \\ &= E[(X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)] \\ &= E[(X - \mu_x)^2] + E[(Y - \mu_y)^2] + 2E[(X - \mu_x)(Y - \mu_y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)\end{aligned}\quad (2.4)$$

Concluimos esta sección mostrando que la varianza de la suma de variables aleatorias independientes es igual a la suma de sus varianzas.

Proposición Si X y Y son variables aleatorias independientes, entonces

$$\text{Cov}(X, Y) = 0$$

y entonces, por la ecuación (2.4),

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Demostración La ecuación (2.3) implica que necesitamos mostrar que $E[XY] = E[X]E[Y]$. Ahora, en el caso discreto,

$$\begin{aligned}E[XY] &= \sum_j \sum_i x_i y_j P\{X = x_i, Y = y_j\} \\ &= \sum_j \sum_i x_i y_j P\{X = x_i\} P\{Y = y_j\} \quad \text{por independencia} \\ &= \sum_j y_j P\{Y = y_j\} \sum_i x_i P\{X = x_i\} \\ &= E[Y]E[X]\end{aligned}$$

Como se cumple un argumento similar en el caso continuo, queda demostrado el resultado. ■

2.7 Desigualdad de Chebyshev y las leyes de los grandes números

Comenzamos con un resultado conocido como desigualdad de Markov.

Proposición: Desigualdad de Markov Si X sólo toma valores no negativos, entonces para cualquier valor $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Demostración Damos una demostración para el caso en que X es una variable aleatoria continua con función de densidad f .

$$\begin{aligned}E[X] &= \int_0^{\infty} xf(x) dx \\ &= \int_0^a xf(x) dx + \int_a^{\infty} xf(x) dx \\ &\geq \int_a^{\infty} xf(x) dx \\ &\geq \int_a^{\infty} af(x) dx \quad \text{pues } xf(x) \geq af(x) \text{ cuando } x \geq a \\ &= a \int_a^{\infty} f(x) dx = aP\{X \geq a\}\end{aligned}$$

y el resultado queda demostrado. ■

Como corolario tenemos la llamada desigualdad de Chebyshev, la cual establece que la probabilidad de que una variable aleatoria difiera de su media por más de k desviaciones estándar está acotada por $1/k^2$, donde la desviación estándar de una variable aleatoria se define como la raíz cuadrada de su varianza.

Corolario: Desigualdad de Chebyshev Si X es una variable aleatoria con media μ y varianza σ^2 , entonces para cualquier valor $k > 0$,

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Demostración Como $(X - \mu)^2/\sigma^2$ es una variable aleatoria no negativa cuya media es

$$E\left[\frac{(X - \mu)^2}{\sigma^2}\right] = \frac{E[(X - \mu)^2]}{\sigma^2} = 1$$

de la desigualdad de Markov obtenemos que

$$P\left\{\frac{(X - \mu)^2}{\sigma^2} \geq k^2\right\} \leq \frac{1}{k^2}$$

De donde deducimos el resultado, pues la desigualdad $(X - \mu)^2/\sigma^2 \geq k^2$ es equivalente a la desigualdad $|X - \mu| \geq k\sigma$. ■

Ahora nos servimos de la desigualdad de Chebyshev para demostrar la ley débil de los grandes números, la cual establece que la probabilidad de que el promedio de los primeros n términos de una sucesión de variables aleatorias independientes e idénticamente distribuidas difiera de su media por más de ϵ tiende a 0 cuando n tiende a infinito.

Teorema: La ley débil de los grandes números Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ . Entonces, para cada $\epsilon > 0$,

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \epsilon\right\} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty$$

Demostración Damos una demostración con la hipótesis adicional de que las variables aleatorias X_i tienen una varianza finita σ^2 . Ahora,

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} (E[X_1] + \dots + E[X_n]) = \mu$$

y

$$\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} [\text{Var}(X_1) + \dots + \text{Var}(X_n)] = \frac{\sigma^2}{n}$$

donde la ecuación anterior utiliza el hecho de que la varianza de la suma de variables aleatorias independientes es igual a la suma de sus varianzas. Por lo tanto, por la desigualdad de Chebyshev, se sigue que para cualquier k positivo

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right\} \leq \frac{1}{k^2}$$

Por lo tanto, para cualquier $\epsilon > 0$, si k es tal que $k\sigma/\sqrt{n} = \epsilon$, es decir, si $k^2 = n\epsilon^2/\sigma^2$, vemos que

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

lo cual establece el resultado. ■

Una generalización de la ley débil es la ley fuerte de los grandes números, la cual establece que, con probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu$$

Es decir, con toda certeza, a largo plazo el promedio de una sucesión de variables independientes e idénticamente distribuidas convergerá a su media.

2.8 Algunas variables aleatorias discretas

Hay ciertas variables aleatorias que surgen con frecuencia en las aplicaciones. En esta sección daremos un panorama de algunas de las discretas.

VARIABLES ALEATORIAS BINOMIALES

Supongamos que se realizan n ensayos independientes, cada uno de los cuales produce un "éxito" con probabilidad p . Si X representa el número de éxitos que ocurren en los n ensayos, entonces X es una variable aleatoria binomial con parámetros (n, p) . Su función de masa de probabilidad está dada por

$$p_i = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n \quad (2.5)$$

donde $\binom{n}{i} = n!/[(n-i)!i!]$ es el coeficiente binomial, igual al número de subconjuntos distintos de i elementos que se pueden elegir de un conjunto de n elementos.

La validez de la ecuación (2.5) se aprecia si comenzamos por observar que la probabilidad de cualquier sucesión particular de resultados que produzca i éxitos y $n - i$ fracasos es, por la supuesta independencia de los ensayos, $p^i(1 - p)^{n-i}$. La ecuación (2.5) es entonces una consecuencia de esto, pues existen $\binom{n}{i}$ sucesiones distintas de los n resultados que producen i éxitos y $n - i$ fracasos, lo cual puede verse al advertir que existen $\binom{n}{i}$ elecciones distintas de los i ensayos que producen éxitos.

Una variable aleatoria binomial $(1, p)$ es una variable aleatoria Bernoulli.

Como, X , variable aleatoria binomial (n, p) representa el número de éxitos en n ensayos independientes, de los que cada uno tiene éxito con probabilidad p , podemos representarla como sigue:

$$X = \sum_{i=1}^n X_i \quad (2.6)$$

donde

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo ensayo es un éxito} \\ 0 & \text{en caso contrario} \end{cases}$$

Ahora,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} = p \\ \text{Var}(X_i) &= E[X_i^2] - E[X_i]^2 \\ &= p - p^2 = p(1 - p) \end{aligned}$$

donde la ecuación anterior utiliza el hecho de que $X_i^2 = X_i$ (pues $0^2 = 0$ y $1^2 = 1$). Por lo tanto, la representación (2.6) implica que, para X variable aleatoria binomial (n, p) ,

$$\begin{aligned} E[X] &= \sum_{i=1}^n E[X_i] = np \\ \text{Var}(X) &= \sum_{i=1}^n \text{Var}(X_i) \quad \text{pues las } X_i \text{ son independientes} \\ &= np(1 - p) \end{aligned}$$

La siguiente fórmula recursiva para p_{i+1} en términos de p_i es útil al calcular las probabilidades binomiales:

$$\begin{aligned} p_{i+1} &= \frac{n!}{(n-i-1)!(i+1)!} p^{i+1}(1-p)^{n-i-1} \\ &= \frac{n!(n-i)}{(n-i)!i!(i+1)} p^i(1-p)^{n-i} \frac{p}{1-p} \\ &= \frac{n-i}{i+1} \frac{p}{1-p} p_i \end{aligned}$$

VARIABLES ALEATORIAS POISSON

Una variable aleatoria X que toma uno de los valores $0, 1, 2, \dots$ es una variable aleatoria Poisson con parámetro λ , $\lambda > 0$, si su función de masa de probabilidad está dada por

$$p_i = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, \dots$$

El símbolo e , definido por $e = \lim_{n \rightarrow \infty} (1 + 1/n)^n$, es una constante famosa en matemáticas, que es aproximadamente igual a 2.7183.

Las variables aleatorias Poisson tienen una gama amplia de aplicaciones, y una razón para ello es que sirven para aproximar la distribución del número de aciertos en muchos ensayos (que sean independientes o, a lo más, "débilmente dependientes") cuando cada uno tiene una pequeña probabilidad de ser un éxito. Para ver por qué, supongamos que X es una variable aleatoria binomial con parámetros (n, p) , de modo que representa el número de aciertos en n ensayos independientes cuando cada uno es un éxito con probabilidad p , y sea $\lambda = np$. Entonces

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)!i!} p^i(1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1) \cdots (n-i+1)}{n^i} \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^{n-i} \end{aligned}$$

Ahora, para n grande y p pequeña,

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1) \cdots (n-i+1)}{n^i} \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^i \approx 1$$

Por lo tanto, para n grande y p pequeña,

$$P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Como la media y la varianza de una variable aleatoria binomial Y están dadas por

$$E[Y] = np, \quad \text{Var}(Y) = np(1-p) \approx np \quad \text{para } p \text{ pequeña}$$

es claro intuitivamente, dada la relación entre las variables aleatorias binomial y Poisson, que para una variable aleatoria Poisson X con parámetro λ ,

$$E[X] = \text{Var}(X) = \lambda$$

Dejamos como ejercicio una demostración analítica de lo anterior.

Para calcular las probabilidades Poisson, hacemos uso de la siguiente fórmula recursiva:

$$\frac{p_{i+1}}{p_i} = \frac{\frac{e^{-\lambda} \lambda^{i+1}}{(i+1)!}}{\frac{e^{-\lambda} \lambda^i}{i!}} = \frac{\lambda}{i+1}$$

o en forma equivalente,

$$p_{i+1} = \frac{\lambda}{i+1} p_i \quad i \geq 0$$

VARIABLES ALEATORIAS GEOMÉTRICAS

Consideremos varios ensayos independientes, de los que cada uno es un éxito con probabilidad p . Si X representa el número del primer ensayo que es un éxito, entonces

$$P\{X = n\} = p(1-p)^{n-1}, \quad n \geq 1 \quad (2.7)$$

lo cual se obtiene fácilmente observando que para que ocurra el primer éxito en el n -ésimo ensayo, los primeros $n-1$ deben ser fracasos y el n -ésimo un éxito. Así, la ecuación (2.7) se deduce pues los ensayos son independientes.

La variable aleatoria cuya función de masa de probabilidad está dada por (2.7) es geométrica con parámetro p . La media de la geométrica se obtiene como sigue:

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p}$$

la ecuación anterior hizo uso de la identidad algebraica, para $0 < x < 1$,

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{d}{dx} \left(\sum_{n=0}^{\infty} x^n \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}$$

Tampoco es difícil mostrar que

$$\text{Var}(X) = \frac{1-p}{p^2}$$

LA VARIABLE ALEATORIA BINOMIAL NEGATIVA

Si X denota el número de ensayos necesarios para acumular un total de r éxitos cuando cada ensayo es, de manera independiente, un éxito con probabilidad p , entonces X es una variable aleatoria binomial negativa, también llamada Pascal, con parámetros p y r . La función de masa de probabilidad de tal variable aleatoria está dada por lo siguiente:

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geq r \quad (2.8)$$

Para ver por qué es válida la ecuación (2.8), observe que para que exactamente n ensayos acumulen r éxitos, los primeros $n-1$ ensayos deben producir exactamente $r-1$ éxitos (la probabilidad de esto es $\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$) y luego el n -ésimo ensayo debe ser un éxito (la probabilidad de esto es p).

Si X_i , $i = 1, \dots, r$, denota el número de ensayos necesarios después del $(i-1)$ -ésimo éxito para obtener el i -ésimo éxito, entonces es fácil ver que éstas son variables aleatorias geométricas independientes con parámetro común p . Como

$$X = \sum_{i=1}^r X_i$$

vemos que

$$E[X] = \sum_{i=1}^r E[X_i] = \frac{r}{p}$$

$$\text{Var}(X) = \sum_{i=1}^r \text{Var}(X_i) = \frac{r(1-p)}{p^2}$$

donde hemos utilizado el resultado correspondiente para variables aleatorias geométricas.

VARIABLES ALEATORIAS HIPERGEOMÉTRICAS

Consideremos una urna con $N+M$ bolas, de las cuales N tienen color claro y M color oscuro. Si se elige una muestra de tamaño n de manera aleatoria [en el sentido de que cada uno de los $\binom{N+M}{n}$ subconjuntos de tamaño n tiene la misma probabilidad de ser elegido], entonces X , el número de bolas de color claro elegidas, tiene la función de masa de probabilidad

$$P\{X = i\} = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}$$

Una variable aleatoria X cuya función de masa de probabilidad está dada por la ecuación anterior es una variable aleatoria hipergeométrica.

Supongamos que las n bolas se eligen de manera secuencial. Sea

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ésima selección es clara} \\ 0 & \text{en caso contrario} \end{cases}$$

entonces

$$X = \sum_{i=1}^n X_i \quad (2.9)$$

y entonces

$$E[X] = \sum_{i=1}^n E[X_i] = \frac{nN}{N+M}$$

donde la ecuación anterior utiliza el hecho de que, por simetría, en la i -ésima selección es igualmente probable obtener cualesquiera de las $N+M$ bolas, por lo que $E[X_i] = P\{X_i = 1\} = N/(N+M)$.

Como las X_i no son independientes (¿por qué?), el uso de la representación (2.9) para calcular $\text{Var}(X)$ implica términos de covarianza. Se puede mostrar que el resultado final es

$$\text{Var}(X) = \frac{nNM}{(N+M)^2} \left(1 - \frac{n-1}{N+M-1} \right)$$

2.9 Variables aleatorias continuas

En esta sección consideramos ciertos tipos de variables aleatorias continuas.

VARIABLES ALEATORIAS DISTRIBUIDAS UNIFORMEMENTE

Una variable aleatoria X está distribuida uniformemente en el intervalo (a, b) , $a < b$, si su función de densidad de probabilidad está dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{sea } a < x < b \\ 0 & \text{en caso contrario} \end{cases}$$

En otras palabras, X está distribuida uniformemente en (a, b) si coloca toda su masa en ese intervalo y es igualmente probable que esté "cerca" de cualquier punto en ese intervalo.

La media y varianza de una variable aleatoria uniforme en (a, b) se obtienen como sigue:

$$E[X] = \frac{1}{b-a} \int_a^b x \, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$E[X^2] = \frac{1}{b-a} \int_a^b x^2 \, dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

y entonces

$$\text{Var}(X) = \frac{1}{3} (a^2 + b^2 + ab) - \frac{1}{4} (a^2 + b^2 + 2ab) = \frac{1}{12} (b-a)^2$$

Así, por ejemplo, el valor esperado es, como es de suponer, el punto medio del intervalo (a, b) .

La función de distribución de X está dada, para $a < x < b$, por

$$F(x) = P\{X \leq x\} = \int_a^x (b-a)^{-1} \, dx = \frac{x-a}{b-a}$$

VARIABLES ALEATORIAS NORMALES

Una variable aleatoria X está distribuida normalmente con media μ y varianza σ^2 si su función de densidad de probabilidad está dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

La densidad normal es una curva en forma de campana que es simétrica con respecto de μ (véase la figura 2.1).

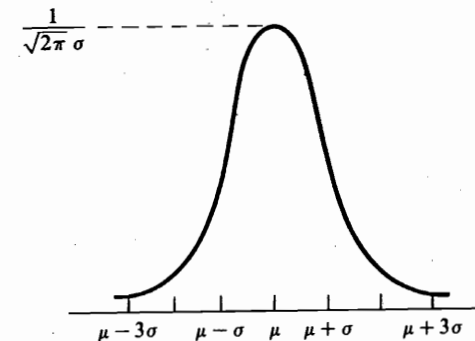


Figura 2.1 La función de densidad normal

No es difícil mostrar que los parámetros μ y σ^2 son iguales a la esperanza y la varianza de la normal. Es decir,

$$E[X] = \mu \quad \text{y} \quad \text{Var}(X) = \sigma^2$$

Un hecho importante en relación con las variables aleatorias normales es que si X es normal con media μ y varianza σ^2 , entonces para cualesquiera constantes a y b , $aX + b$ está distribuida normalmente con media $a\mu + b$ y varianza $a^2\sigma^2$. Esto implica que si X es normal con media μ y varianza σ^2 , entonces

$$Z = \frac{X - \mu}{\sigma}$$

es normal con media 0 y varianza 1. Tal variable aleatoria Z tiene una distribución normal estándar (o unitaria). Sea ϕ la función de distribución de una variable aleatoria normal estándar; es decir

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx, \quad -\infty < x < \infty$$

El resultado de que $Z = (X - \mu)/\sigma$ tiene una distribución normal estándar cuando X es normal con media μ y varianza σ^2 es bastante útil, pues nos permite evaluar todas las probabilidades relativas a X en términos de ϕ . Por ejemplo, la función de distribución de X se puede expresar como

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} \\ &= P\left\{Z \leq \frac{x - \mu}{\sigma}\right\} \\ &= \phi\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

El valor de $\phi(x)$ se puede determinar por medio de una tabla o escribiendo un programa de computadora para aproximarlos.

Para a en el intervalo $(0, 1)$, sea z_a tal que

$$P\{Z > z_a\} = 1 - \phi(z_a) = a$$

Es decir, una normal estándar excederá a z_a con probabilidad a (véase la figura 2.2). Los valores de z_a se pueden obtener mediante una tabla de los valores de ϕ . Por ejemplo, como

$$\phi(1.64) = 0.95, \quad \phi(1.96) = 0.975, \quad \phi(2.33) = 0.99$$

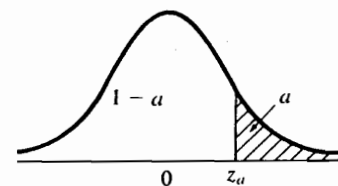


Figura 2.2 $P\{Z > z_a\} = a$

vemos que

$$z_{.05} = 1.64, \quad z_{.025} = 1.96, \quad z_{.01} = 2.33$$

La amplia gama de aplicaciones de las variables aleatorias normales surge de uno de los teoremas más importantes de la teoría de la probabilidad, el teorema central del límite, el cual afirma que la suma de un gran número de variables aleatorias independientes tiene aproximadamente una distribución normal. La forma más sencilla de este importante teorema es la siguiente.

El teorema central del límite Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con media finita μ y varianza finita σ^2 . Entonces

$$\lim_{n \rightarrow \infty} P\left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x\right\} = \phi(x)$$

VARIABLES ALEATORIAS EXPONENCIALES

Una variable aleatoria continua con función de densidad de probabilidad

$$f(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty$$

para cierta $\lambda > 0$ es una variable aleatoria exponencial con parámetro λ . Su distribución acumulativa está dada por

$$F(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}, \quad 0 < x < \infty$$

Es fácil verificar que su valor esperado y su varianza son los siguientes:

$$E[X] = \frac{1}{\lambda} \quad \text{y} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

La propiedad clave de las variables aleatorias exponenciales es que “no tienen memoria”; decimos que la variable aleatoria no negativa X no tiene memoria si

$$P\{X > s + t | X > s\} = P\{X > t\} \quad \text{para toda } s, t \geq 0 \quad (2.10)$$

Para entender por qué se dice esto, imagine que X representa el tiempo de vida de cierta unidad y considere la probabilidad de que una unidad de edad s sobreviva un tiempo adicional t . Como esto ocurrirá si el tiempo de vida de la unidad excede $t + s$ dado que sigue viva en el instante s , vemos que

$$P\{\text{la vida adicional de un elemento de edad } s \text{ excede a } t\} = P\{X > s + t | X > s\}$$

Así, la ecuación (2.10) enuncia el hecho de que la distribución de la vida restante de un elemento de edad s no dependa de s . Es decir, no es necesario recordar la edad de una unidad para conocer la distribución de su vida restante.

La ecuación (2.10) es equivalente a

$$P\{X > s + t\} = P\{X > s\} P\{X > t\}$$

Como la ecuación anterior se satisface siempre que X sea una variable aleatoria exponencial, pues en este caso $P\{X > x\} = e^{-\lambda x}$, vemos que las variables aleatorias exponenciales no tienen memoria; de hecho, no es difícil mostrar que son las únicas variables aleatorias sin memoria.

Otra propiedad útil de las variables aleatorias exponenciales es que siguen siendo exponenciales al multiplicarse por una constante positiva. Para verlo, supongamos que X es exponencial con parámetro λ , y sea c un número positivo. Entonces

$$P\{cX \leq x\} = P\left\{X \leq \frac{x}{c}\right\} = 1 - e^{-\lambda x/c}$$

lo cual muestra que cX es exponencial con parámetro λ/c .

EL PROCESO POISSON Y LAS VARIABLES ALEATORIAS GAMMA

Supongamos que ocurren ciertos “eventos” en instantes aleatorios y sea $N(t)$ el número de eventos que ocurren en el intervalo de tiempo $[0, t]$. Estos eventos constituyen un *proceso Poisson con razón λ , $\lambda > 0$* , si

- (a) $N(0) = 0$.
- (b) El número de eventos que ocurren en intervalos de tiempo distintos son independientes.
- (c) La distribución del número de eventos que ocurren en un intervalo dado depende solamente de la longitud del intervalo y no de su posición.

$$(d) \lim_{h \rightarrow 0} \frac{P\{N(h) = 1\}}{h} = \lambda.$$

$$(e) \lim_{h \rightarrow 0} \frac{P\{N(h) \geq 2\}}{h} = 0.$$

Así, la condición (a) establece que el proceso comienza en el instante 0. La condición (b), la hipótesis de *incremento independiente*, establece que el número de eventos hasta el instante t [i.e., $N(t)$] es independiente del número de eventos ocurridos entre t y $t + s$ [i.e., $N(t + s) - N(t)$]. La condición (c), la hipótesis de *incremento estacionario*, dice que la distribución de probabilidad de $N(t + s) - N(t)$ es la misma para todos los valores de t . Las condiciones (d) y (e) establecen que en un pequeño intervalo de longitud h , la probabilidad de que ocurra un evento es aproximadamente λh , mientras que la probabilidad de dos o más es aproximadamente 0.

Ahora argumentaremos que estas hipótesis implican que el número de eventos que ocurren en un intervalo de longitud t es una variable aleatoria Poisson con media λt . Para ello, consideremos el intervalo $[0, t]$ y dividámoslo en n subintervalos no traslapados de longitud t/n (figura 2.3). Primero consideremos el número de estos subintervalos que contienen un evento. Como cada subintervalo contiene de manera independiente [por la condición (b)] un evento con la misma probabilidad [por la condición (c)], lo cual es aproximadamente igual a $\lambda t/n$, esto implica que el número de tales intervalos es una variable aleatoria binomial con parámetros n y $p \approx \lambda t/n$. Por lo tanto, por el argumento que asegura la convergencia de la binomial a la Poisson, al hacer $n \rightarrow \infty$ vemos que el número de tales subintervalos converge a una variable aleatoria Poisson con media λt . Como se puede mostrar que la condición (e) implica que la probabilidad de que cualquiera de estos subintervalos contenga dos o más eventos tiende a 0 cuando $n \rightarrow \infty$, tenemos que $N(t)$, el número de eventos que ocurren en $[0, t]$ es una variable aleatoria Poisson con media λt .

Para un proceso Poisson, denotamos por X_1 el tiempo del primer evento. Además, para $n > 1$, sea X_n el tiempo transcurrido entre el $(n - 1)$ -ésimo y el n -ésimo evento. La sucesión $\{X_n, n = 1, 2, \dots\}$ es la *sucesión de tiempos entre llegadas*. Por ejemplo, si $X_1 = 5$ y $X_2 = 10$, entonces el primer evento del proceso Poisson ocurrirá en el instante 5 y el segundo en el instante 15.



Figura 2.3

Ahora determinamos la distribución de las X_n . Primero, observemos que el evento $\{X_1 > t\}$ ocurre si y sólo si ninguno de los eventos del proceso Poisson ocurre en el intervalo $[0, t]$, y entonces

$$P\{X_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

Por lo tanto, X_1 tiene una distribución exponencial con media $1/\lambda$. Para obtener la distribución de X_2 , observemos que

$$\begin{aligned} P\{X_2 > t \mid X_1 = s\} &= P\{0 \text{ eventos en } (s, s+t] \mid X_1 = s\} \\ &= P\{0 \text{ eventos en } (s, s+t]\} \\ &= e^{-\lambda t} \end{aligned}$$

donde las dos últimas ecuaciones son consecuencia de los incrementos independientes y estacionarios. Por lo tanto, de lo anterior concluimos que X_2 también es una variable aleatoria exponencial con media $1/\lambda$, y además, que X_2 es independiente de X_1 . Al repetir el mismo argumento varias veces obtenemos:

Proposición Los tiempos entre llegadas X_1, X_2, \dots son variables aleatorias exponenciales independientes e idénticamente distribuidas con parámetro λ .

Sea $S_n = \sum_{i=1}^n X_i$ el tiempo del n -ésimo evento. Como S_n será menor o igual que t si y sólo si han ocurrido al menos n eventos hasta el instante t , vemos que

$$\begin{aligned} P\{S_n \leq t\} &= P\{N(t) \geq n\} \\ &= \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} \end{aligned}$$

Como el lado izquierdo es la función de distribución acumulativa de S_n , obtenemos, por derivación, que la función de densidad de S_n , denotada $f_n(t)$, está dada por

$$\begin{aligned} f_n(t) &= \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

Definición Una variable aleatoria con función de densidad de probabilidad

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t > 0$$

es una variable aleatoria gamma con parámetros (n, λ) .

Así, vemos que S_n , el tiempo del n -ésimo evento de un proceso Poisson con razón λ , es una variable aleatoria gamma con parámetros (n, λ) . Además, de la representación $S_n = \sum_{i=1}^n X_i$ y de la proposición anterior, la cual establece que estas X_i son exponenciales independientes con razón λ , tenemos el siguiente corolario.

Corolario La suma de n variables aleatorias exponenciales independientes, cada una con parámetro λ , es una variable aleatoria gamma con parámetros (n, λ) .

EL PROCESO POISSON NO HOMOGÉNEO

Desde el punto de vista de la modelación, la principal debilidad del proceso Poisson es la hipótesis de que los eventos tienen la misma probabilidad de ocurrir en todos los intervalos del mismo tamaño. Una generalización, que relaja esta hipótesis, conduce al llamado proceso no homogéneo o no estacionario.

Si los "eventos" ocurren de manera aleatoria en el transcurso del tiempo, y $N(t)$ denota el número de ellos que ocurre en el instante t , entonces decimos que $\{N(t), t \geq 0\}$ constituye un proceso Poisson no homogéneo con función de intensidad $\lambda(t)$, $t \geq 0$, si

- $N(0) = 0$.
- Los números de eventos que ocurren en intervalos de tiempo ajenos son independientes.
- $\lim_{h \rightarrow 0} P\{\text{exactamente 1 evento entre } t \text{ y } t+h\}/h = \lambda(t)$.
- $\lim_{h \rightarrow 0} P\{\text{dos o más eventos entre } t \text{ y } t+h\}/h = 0$.

La función $m(t)$, definida como

$$m(t) = \int_0^t \lambda(s) ds, \quad t \geq 0$$

es la función de valor medio. Se puede establecer el siguiente resultado.

Proposición $N(t+s) - N(t)$ es una variable aleatoria Poisson con media $m(t+s) - m(t)$.

La cantidad $\lambda(t)$, llamada la intensidad en el instante t , indica la probabilidad de que un evento ocurra cerca del instante t [observe que cuando $\lambda(t) \equiv \lambda$, el proceso no homogéneo se convierte en el proceso Poisson usual]. La siguiente proposición indica una forma útil de interpretar un proceso Poisson no homogéneo.

Proposición Supongamos que ciertos eventos ocurren de acuerdo con un proceso Poisson con razón λ , y supongamos también que, independientemente de lo que haya pasado antes, un evento que ocurre en el instante t se cuenta con probabilidad $p(t)$. Entonces, el proceso de los eventos contados constituye un proceso Poisson no homogéneo con función de intensidad $\lambda(t) = \lambda p(t)$.

Demostración Esta proposición se demuestra observando que se satisfacen todas las condiciones dadas anteriormente. Las condiciones (a), (b) y (d) son consecuencia del hecho de que el resultado correspondiente es verdadero para todos los eventos (no sólo para los eventos contados). Las condiciones (c) son consecuencia de

$$\begin{aligned} P\{1 \text{ evento contado entre } t \text{ y } t+h\} \\ &= P\{1 \text{ evento y es contado}\} \\ &+ P\{2 \text{ o más eventos y exactamente 1 es contado}\} \\ &\approx \lambda h p(t) \end{aligned}$$

2.10 Esperanza condicional y varianza condicional

Si X y Y son variables aleatorias conjuntamente discretas, definimos $E[X|Y=y]$, la esperanza condicional de X dado que $Y=y$, como

$$\begin{aligned} E[X|Y=y] &= \sum_x xP\{X=x|Y=y\} \\ &= \frac{\sum_x xP\{X=x, Y=y\}}{P\{Y=y\}} \end{aligned}$$

En otras palabras, la esperanza condicional de X , dado que $Y=y$, se define como antes, como un promedio ponderado de todos los valores posibles de X , pero ahora el peso dado al valor x es igual a la probabilidad condicional de que X sea igual a x dado que Y es igual a y .

De manera similar, si X y Y son conjuntamente continuas con función de densidad conjunta $f(x, y)$, definimos la esperanza condicional de X , dado que $Y=y$, como

$$E[X|Y=y] = \frac{\int xf(x, y) dx}{\int f(x, y) dx}$$

Sea $E[X|Y]$ la función de la variable aleatoria Y cuyo valor en $Y=y$ es $E[X|Y=y]$; observe que $E[X|Y]$ es en sí una variable aleatoria. La siguiente proposición es bastante útil.

Proposición

$$E[E[X|Y]] = E[X] \quad (2.11)$$

Si Y es una variable aleatoria discreta, entonces la ecuación (2.11) establece que

$$E[X] = \sum_y E[X|Y=y]P\{Y=y\}$$

mientras que si Y es continua con densidad g , entonces (2.11) establece que

$$E[X] = \int E[X|Y=y]g(y) dy$$

Ahora daremos una demostración de la proposición anterior cuando X y Y son discretas:

$$\begin{aligned} \sum_y E[X|Y=y]P\{Y=y\} &= \sum_y \sum_x xP\{X=x|Y=y\}P\{Y=y\} \\ &= \sum_y \sum_x xP\{X=x, Y=y\} \\ &= \sum_x x \sum_y P\{X=x, Y=y\} \\ &= \sum_x xP\{X=x\} \\ &= E[X] \end{aligned}$$

También podemos definir la varianza condicional de X , dado el valor de Y , como sigue:

$$\text{Var}(X|Y) = E[(X - E[X|Y])^2|Y]$$

Es decir, $\text{Var}(X|Y)$ es una función de Y , que en $Y=y$ es igual a la varianza de X dado que $Y=y$. Mediante el mismo razonamiento que nos condujo a la identidad $\text{Var}(X) = E[X^2] - (E[X])^2$ tenemos que

$$\text{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

Al calcular las esperanzas de ambos lados de la ecuación tenemos

$$\begin{aligned} E[\text{Var}(X|Y)] &= E[E[X^2|Y]] - E[(E[X|Y])^2] \\ &= E[X^2] - E[(E[X|Y])^2] \end{aligned} \quad (2.12)$$

Además, como $E[E[X|Y]] = E[X]$, tenemos que

$$\text{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[X])^2 \quad (2.13)$$

Al sumar las ecuaciones (2.12) y (2.13), obtenemos la siguiente identidad, conocida como la fórmula de la varianza condicional.

La fórmula de la varianza condicional

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

Ejercicios

1. Considere un experimento que consta de seis caballos, numerados del 1 al 6, que realizan una carrera, y suponga que el espacio muestral está dado por

$$S = \{\text{todos los ordenamientos de } (1, 2, 3, 4, 5, 6)\}$$

Sea A el evento de que el caballo número 1 esté entre los tres primeros finalistas, sea B el evento de que el caballo número 2 llegue en segundo, y sea C el evento de que el caballo número 3 llegue en tercero.

- Describa el evento $A \cup B$. ¿Cuántos resultados están contenidos en este evento?
- ¿Cuántos resultados están contenidos en el evento AB ?
- ¿Cuántos resultados están contenidos en el evento ABC ?
- ¿Cuántos resultados están contenidos en el evento $A \cup BC$?

2. (a) Para cualesquiera eventos A y B , muestre que

$$A \cup B = A \cup A^c B$$

$$A = AB \cup AB^c$$

- (b) Muestre que

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

- Una pareja tiene dos hijos. ¿Cuál es la probabilidad de que ambos sean mujeres, dado que la mayor es una mujer? Suponga que las cuatro posibilidades son igualmente probables.
- El rey proviene de una familia de dos hijos. ¿Cuál es la probabilidad de que el otro hijo sea hombre?
- La variable aleatoria X toma uno de los valores 1, 2, 3, 4 con probabilidades

$$P\{X = i\} = ic, \quad i = 1, 2, 3, 4$$

para algún valor c . Determine $P\{2 \leq X \leq 3\}$.

- La variable aleatoria continua X tiene una función de densidad de probabilidad dada por

$$f(x) = cx, \quad 0 < x < 1$$

Determine $P\{X > \frac{1}{2}\}$.

- Si X y Y tienen una función de densidad de probabilidad conjunta dada por

$$f(x, y) = 2e^{-(x+2y)}, \quad 0 < x < \infty, 0 < y < \infty$$

determine $P\{X < Y\}$.

- Determine el valor esperado de la variable aleatoria dada en el ejercicio 5.
- Determine $E[X]$ para la variable aleatoria del ejercicio 6.
- Existen 10 tipos distintos de cupones y cada vez que alguien recibe un cupón, es igualmente probable que sea cualquiera de los 10 tipos. Sea X el número de

tipos distintos contenidos en una colección de N cupones. Determine $E[X]$. [Sugerencia: Para $i = 1, \dots, 10$ sea

$$X_i = \begin{cases} 1 & \text{si un cupón de tipo } i \text{ está entre los } N \\ 0 & \text{en caso contrario} \end{cases}$$

y haga uso de la representación $X = \sum_{i=1}^{10} X_i$]

- Se lanza un dado con seis lados. Si cada uno de los seis resultados posibles es igualmente probable, determine la varianza del número que aparece hacia arriba.
- Suponga que X tiene la función de densidad de probabilidad

$$f(x) = ce^x, \quad 0 < x < 1$$

Determine $\text{Var}(X)$.

- Muestre que $\text{Var}(aX + b) = a^2 \text{Var}(X)$.
- Suponga que X , la cantidad de jugo de manzana contenido en un recipiente comercial, es una variable aleatoria con media de 4 gramos.
 - ¿Qué se puede decir acerca de la probabilidad de que un recipiente dado contenga más de 6 gramos de jugo de manzana?
 - Si $\text{Var}(X) = 4(\text{gramos})^2$, ¿qué se puede decir acerca de la probabilidad de que un recipiente dado contenga entre 3 y 5 gramos de jugo de manzana?
- Un avión necesita al menos la mitad de sus motores para concluir su misión de manera segura. Si cada motor funciona de manera independiente con probabilidad p , ¿para qué valores de p es un avión trimotor más seguro que un avión con cinco motores?
- Para una variable aleatoria binomial X con parámetros (n, p) , muestre que $P\{X = i\}$ aumenta primero y luego disminuye, alcanzando su valor máximo cuando i es el máximo entero menor o igual a $(n + 1)p$.
- Si X y Y son variables aleatorias binomiales independientes con parámetros respectivos (n, p) y (m, p) , argumente, sin realizar cálculos, que $X + Y$ es binomial con parámetros $(n + m, p)$.
- Explique por qué todas las variables aleatorias siguientes tienen aproximadamente una distribución Poisson:
 - El número de errores tipográficos en un capítulo dado de este libro.
 - El número de números telefónicos marcados en forma incorrecta cada día.
 - El número de clientes que entran a una oficina de correos un día dado.
- Si X es una variable aleatoria Poisson con parámetro λ , muestre que
 - $E[X] = \lambda$.
 - $\text{Var}(X) = \lambda$.

20. Sean X y Y variables aleatorias Poisson independientes con parámetros λ_1 y λ_2 , respectivamente. Utilice el resultado del ejercicio 17 para argumentar de manera heurística que $X + Y$ es Poisson con parámetro $\lambda_1 + \lambda_2$. Después dé una demostración analítica de esto. [Sugerencia:

$$P\{X + Y = k\} = \sum_{i=0}^k P\{X = i, Y = k - i\} = \sum_{i=0}^k P\{X = i\}P\{Y = k - i\}$$

21. Explique cómo utilizar la relación

$$p_{i+1} = \frac{\lambda}{i+1} p_i$$

para calcular de manera eficiente las probabilidades Poisson.

22. Determine $P\{X > n\}$ cuando X es una variable aleatoria geométrica con parámetro p .
23. Dos jugadoras practican cierto juego, hasta que una de ellas ha ganado un total de cinco partidas. Si la jugadora A gana cada partida con probabilidad 0.6, ¿cuál es la probabilidad de que gane el encuentro?
24. Considere el modelo hipergeométrico de la sección 2.8 y suponga que las bolas claras están numeradas. Para $i = 1, \dots, N$, sea

$$Y_i = \begin{cases} 1 & \text{si se elige la bola clara con el número } i \\ 0 & \text{en caso contrario} \end{cases}$$

Argumente que $X = \sum_{i=1}^N Y_i$, y luego emplee esta representación para determinar $E[X]$. Verifique esto con el resultado dado en la sección 2.8.

25. El autobús llegará en un instante que tiene una distribución uniforme entre las ocho y las ocho y media de la mañana. Si nos presentamos a las ocho, ¿cuál es la probabilidad de que esperemos entre 5 y 15 minutos?
26. Para una variable aleatoria normal con parámetros μ y σ^2 , muestre que
- (a) $E[X] = \mu$.
- (b) $\text{Var}(X) = \sigma^2$.
27. Sea X una variable aleatoria binomial con parámetros (n, p) . Explique por qué

$$P\left\{\frac{X - np}{\sqrt{np(1-p)}} \leq x\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$$

cuando n es grande.

28. Si X es una variable aleatoria exponencial con parámetro λ , muestre que
- (a) $E[X] = 1/\lambda$.
- (b) $\text{Var}(X) = 1/\lambda^2$.

29. Las personas A , B y C esperan en un banco que tiene dos cajeros al abrir en la mañana. Las personas A y B van hacia un cajero y C espera en la fila. Si el tiempo que tarda un cajero en atender a un cliente es una variable aleatoria exponencial con parámetro λ , ¿cuál es la probabilidad de que C sea el último en salir del banco? [Sugerencia: no se necesitan cálculos.]
30. Sean X y Y variables aleatorias exponenciales independientes, con razones respectivas λ y μ . Muestre que

$$P\{X < Y\} = \frac{\lambda}{\lambda + \mu}$$

31. Considere un proceso Poisson en el cual los eventos ocurren a razón de 0.3 por hora. ¿Cuál es la probabilidad de que ningún evento ocurra entre las 10 de la mañana y las dos de la tarde?
32. Para un proceso Poisson con razón λ , determine $P\{N(s) = k \mid N(t) = n\}$ cuando $s < t$.
33. Repita el ejercicio 32 para $s > t$.
34. Si X es una variable aleatoria gamma con parámetros (n, p) , determine
- (a) $E[X]$.
- (b) $\text{Var}(X)$.
35. Una urna contiene cuatro bolas blancas y seis negras. Se elige una muestra aleatoria de tamaño 4. Sea X el número de bolas blancas en la muestra. Ahora se elige una bola adicional de las seis bolas restantes en la urna. Sea Y igual a 1 si esta bola es blanca y 0 si es negra. Determine
- (a) $E[Y \mid X = 2]$.
- (b) $E[X \mid Y = 1]$.
- (c) $\text{Var}(Y \mid X = 0)$.
- (d) $\text{Var}(X \mid Y = 1)$.
36. Si X y Y son variables aleatorias exponenciales independientes e idénticamente distribuidas, muestre que la distribución condicional de X , dado que $X + Y = t$, es la distribución uniforme en $(0, t)$.

Bibliografía

- Feller, W., *An Introduction to Probability Theory and its Applications*, Nueva York, Wiley, 3a. ed., 1968.
- Ross, S. M., *A First Course in Probability*, Nueva York, Macmillan, 4a. ed., 1993.
- *Introduction to Probability Models*, Nueva York, Academic Press, 5a. ed., 1993.

Introducción

El núcleo de un estudio de simulación es la capacidad de generar números aleatorios, que representan el valor de una variable aleatoria distribuida uniformemente en $(0, 1)$. En este capítulo explicamos la forma de hacerlo mediante una computadora y también comenzamos a ilustrar sus usos.

3.1 Generación de números pseudoaleatorios

Al principio, los números aleatorios se generaban en forma manual o mecánica, utilizando técnicas como ruedas giratorias, lanzamientos de dados o barajas; pero el planteamiento moderno consiste en utilizar una computadora para generar de manera sucesiva números pseudoaleatorios. Estos números constituyen una sucesión de valores que, aunque son producidos de manera determinista, tienen toda la apariencia de ser variables aleatorias uniformes e independientes en $(0, 1)$.

Uno de los métodos más comunes para generar números pseudoaleatorios comienza con un valor inicial x_0 , llamado semilla, y luego se calcula de manera recursiva los valores sucesivos x_n , $n \geq 1$, haciendo

$$x_n = ax_{n-1} \text{ módulo } m \quad (3.1)$$

donde a y m son enteros positivos dados y lo anterior significa que ax_{n-1} se divide entre m y el residuo se considera como el valor de x_n . Así, cada x_n es $0, 1, \dots, m-1$ y la

Sec. 3.1 Generación de números pseudoaleatorios

37

cantidad x_n/m (llamada número pseudoaleatorio) se considera como una aproximación del valor de una variable aleatoria uniforme en $(0, 1)$.

El método especificado por la ecuación (3.1) para generar números aleatorios se llama el método congruencial multiplicativo. Como cada uno de los números x_n asume uno de los valores $0, 1, \dots, m-1$, se tiene que después de cierto número finito (a lo más m) de valores generados, alguno debe repetirse, y, una vez que esto ocurre, toda la sucesión comienza a repetirse. Así, queremos encontrar las constantes a y m tales que, para cualquier semilla inicial x_0 , el número de variables que se puedan generar antes de que suceda esta repetición sea "grande".

En general, las constantes a y m deben satisfacer tres criterios:

1. Para cualquier semilla inicial, la sucesión resultante tiene la "apariencia" de ser una sucesión de variables aleatorias independientes y uniformes en $(0, 1)$.
2. Para cualquier semilla inicial, el número de variables que se pueden generar antes de que comience la repetición es grande.
3. Los valores se pueden calcular de manera eficiente en una computadora digital.

Un criterio que parece ser útil para cumplir las tres condiciones anteriores es que m debe ser un número primo grande, adecuado al tamaño de la palabra de la computadora. Para una máquina con palabras de 32 bits (donde el primero es un bit de signo), se ha mostrado que las elecciones de $m = 2^{31} - 1$ y $a = 7^5 = 16\,807$ producen las propiedades deseadas (para una máquina con palabras de 36 bits, parece que funcionan las elecciones de $m = 2^{35} - 31$ y $a = 5^5$).

Otro generador de números pseudoaleatorios utiliza fórmulas recursivas del tipo

$$x_n = (ax_{n-1} + c) \text{ módulo } m$$

Tales fórmulas se llaman generadores congruenciales mixtos (ya que tienen un término aditivo y uno multiplicativo). Al emplearlos, con frecuencia se elige m como la longitud de la palabra de la computadora, ya que esto hace más eficiente el cálculo de $(ax_{n-1} + c)$ módulo m (es decir, la división de $ax_{n-1} + c$ entre m).

La mayor parte de los lenguajes de computadora tienen integrado un generador de números aleatorios que se puede llamar con tal fin. Por ejemplo, BASIC utiliza dos instrucciones:

- 1 RANDOMIZE
- 2 U = RND

El resultado de la instrucción RANDOMIZE es la solicitud al usuario de que introduzca la semilla x_0 . Una vez hecho esto, la cantidad RND representará el siguiente

número pseudoaleatorio. Por ejemplo, este programa imprime los valores de 10 números pseudoaleatorios.

```
10 RANDOMIZE
20 FOR I = 1 TO 10
30 U = RND
40 PRINT U
50 NEXT
```

Como punto de partida en la simulación de sistemas por computadora, supondremos que podemos generar una sucesión de números pseudoaleatorios que se pueden considerar como una aproximación a los valores de una sucesión de variables aleatorias independientes y uniformes en $(0, 1)$. Es decir, no exploraremos las interesantes cuestiones teóricas, que requieren material más allá de los objetivos de este texto, relacionadas con la construcción de "buenos" generadores de números pseudoaleatorios. Mejor supondremos que tenemos una "caja negra" que proporciona un número aleatorio si así se solicita.

3.2 Uso de números aleatorios para evaluar integrales

Una de las primeras aplicaciones de los números aleatorios fue en el cálculo de integrales. Sea $g(x)$ una función y supongamos que queremos calcular θ , donde

$$\theta = \int_0^1 g(x) dx$$

Para calcular el valor de θ , observe que si U está distribuida uniformemente en $(0, 1)$, entonces podemos expresar θ como

$$\theta = E[g(U)]$$

Si U_1, \dots, U_k son variables aleatorias independientes y uniformes en $(0, 1)$, esto implica que $g(U_1), \dots, g(U_k)$ son variables aleatorias independientes e idénticamente distribuidas con media θ . Por lo tanto, por la ley fuerte de los grandes números, se tiene, con probabilidad 1,

$$\sum_{i=1}^k \frac{g(U_i)}{k} \rightarrow E[g(U)] = \theta \quad \text{cuando } k \rightarrow \infty$$

Por lo tanto, podemos aproximar θ generando un gran número de números aleatorios u_i y considerando como nuestra aproximación a θ el valor promedio de $g(u_i)$. Este método de aproximación de integrales es el método *Monte Carlo*.

Sec. 3.2 Uso de números aleatorios para evaluar integrales

Se puede utilizar el siguiente algoritmo en BASIC para aproximar θ .

```
10 RANDOMIZE
20 INPUT K
30 S = 0
40 FOR I = 1 TO K
50 U = RND
60 S = S + g(U)
70 NEXT
80 PRINT S/K
```

El valor impreso es la aproximación de $\theta = \int_0^1 g(x) dx$.
Si quisiéramos calcular

$$\theta = \int_a^b g(x) dx$$

entonces, al hacer la sustitución $y = (x - a)/(b - a)$, $dy = dx/(b - a)$, vemos que

$$\begin{aligned} \theta &= \int_0^1 g(a + [b - a]y) (b - a) dy \\ &= \int_0^1 h(y) dy \end{aligned}$$

donde $h(y) = (b - a)g(a + [b - a]y)$. Así, podemos aproximar θ al generar de manera continua números aleatorios y luego considerar el valor promedio de h evaluada en estos números aleatorios.

De manera análoga, si queremos el valor de

$$\theta = \int_0^\infty g(x) dx$$

podríamos aplicar la sustitución $y = 1/(x + 1)$, $dy = -dx/(x + 1)^2 = -y^2 dx$, para obtener la identidad

$$\theta = \int_0^1 h(y) dy$$

donde

$$h(y) = \frac{g\left(\frac{1}{y} - 1\right)}{y^2}$$

La utilidad de los números aleatorios para aproximar integrales es más evidente en el caso de las integrales múltiples. Supongamos que g es una función con un argumento de dimensión n y que estamos interesados en calcular

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

La clave del método Monte Carlo para estimar θ está en el hecho de que θ se puede expresar mediante la siguiente esperanza:

$$\theta = E[g(U_1, \dots, U_n)]$$

donde U_1, \dots, U_n son variables aleatorias independientes y uniformes en $(0, 1)$. Por lo tanto, si generamos k conjuntos independientes, cada uno formado por n variables aleatorias independientes y uniformes en $(0, 1)$

$$U_1^1, \dots, U_n^1$$

$$U_1^2, \dots, U_n^2$$

$$\vdots$$

$$U_1^k, \dots, U_n^k$$

entonces, como las variables aleatorias $g(U_1^i, \dots, U_n^i), i = 1, \dots, k$, son todas variables aleatorias independientes e idénticamente distribuidas con media θ , podemos estimar θ mediante $\sum_{i=1}^k g(U_1^i, \dots, U_n^i)/k$.

Para una aplicación de lo anterior, consideremos el siguiente método para estimar π .

Ejemplo 3a La estimación de π . Supongamos que el vector aleatorio (X, Y) se distribuye uniformemente en el cuadrado de área 4 con centro en el origen. Es decir, es un punto aleatorio en la siguiente región, especificada en la figura 3.1. Consideremos ahora la probabilidad de que este punto aleatorio del cuadrado esté dentro del círculo inscrito de radio 1 (véase la figura 3.2). Observe que, como (X, Y) está distribuido uniformemente en el cuadrado, se tiene que

$$\begin{aligned} P\{(X, Y) \text{ esté en el círculo}\} &= P\{X^2 + Y^2 \leq 1\} \\ &= \frac{\text{Área del círculo}}{\text{Área del cuadrado}} = \frac{\pi}{4} \end{aligned}$$

Por lo tanto, si generamos un gran número de puntos aleatorios en el cuadrado, la proporción de puntos que caen dentro del círculo será aproximadamente $\pi/4$. Ahora,

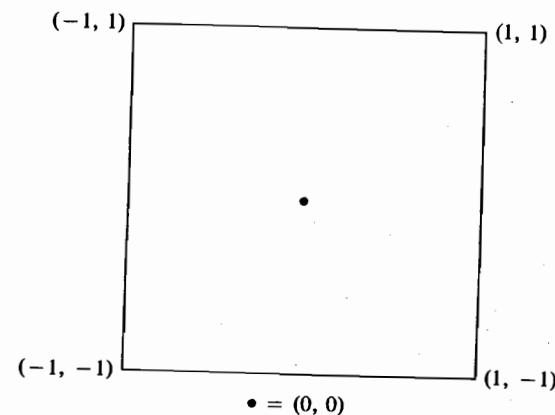


Figura 3.1

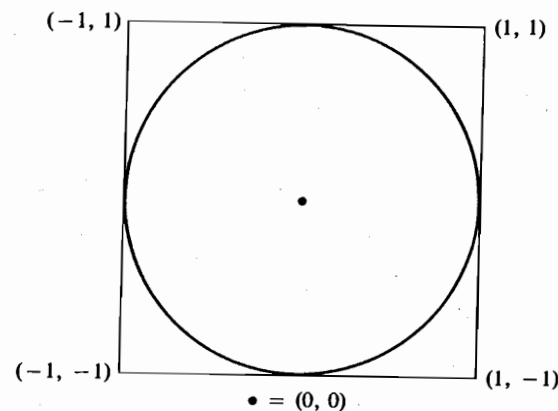


Figura 3.2

si X y Y fuesen independientes y ambas estuviesen uniformemente distribuidas en $(-1, 1)$, su densidad conjunta sería

$$\begin{aligned} f(x, y) &= f(x)f(y) \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4}, \quad -1 \leq x \leq 1, \quad -1 \leq y \leq 1 \end{aligned}$$

Como la función de densidad de (X, Y) es constante en el cuadrado, se tiene (por definición) que (X, Y) está distribuido uniformemente en el cuadrado. Ahora, si U es uniforme en $(0, 1)$, entonces $2U$ es uniforme en $(0, 2)$, por lo cual $2U - 1$ es uniforme en $(-1, 1)$. Por lo tanto, si generamos números aleatorios U_1 y U_2 , hacemos $X = 2U_1 - 1$ y $Y = 2U_2 - 1$ y definimos

$$I = \begin{cases} 1 & \text{si } X^2 + Y^2 \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

entonces

$$E[I] = P\{X^2 + Y^2 \leq 1\} = \frac{\pi}{4}.$$

Por lo tanto, para estimar $\pi/4$ generamos un gran número de pares de números aleatorios u_1, u_2 y estimamos $\pi/4$ como la fracción de pares para los cuales $(2u_1 - 1)^2 + (2u_2 - 1)^2 \leq 1$. ■

Así, con los generadores de números aleatorios se producen los valores de variables aleatorias uniformes en $(0, 1)$. Partiremos de estos números aleatorios para mostrar, en los capítulos 4 y 5, la forma de generar los valores de variables aleatorias a partir de distribuciones arbitrarias. Esta capacidad de generar variables aleatorias arbitrarias nos permitirá simular sistemas de probabilidad; es decir, podremos generar, de acuerdo con las leyes de probabilidad específicas de los sistemas, todas sus cantidades aleatorias conforme transcurra el tiempo.

Ejercicios

1. Si $x_0 = 5$ y

$$x_n = 3x_{n-1} \text{ mód } 150$$

determine x_1, \dots, x_{10} .

2. Si $x_0 = 3$ y

$$x_n = 5x_{n-1} + 7 \text{ mód } 200$$

determine x_1, \dots, x_{10} .

En los ejercicios 3-9, emplee la simulación para aproximar las siguientes integrales. Compare su estimación con la respuesta exacta, si ésta se conoce.

Ejercicios

3. $\int_0^1 \exp\{e^x\} dx$

4. $\int_0^1 (1 - x^2)^{3/2} dx$

5. $\int_{-2}^2 e^{x+x^2} dx$

6. $\int_0^\infty x(1 + x^2)^{-2} dx$

7. $\int_{-\infty}^\infty e^{-x^2} dx$

8. $\int_0^1 \int_0^1 e^{(x+y)^2} dy dx$

9. $\int_0^\infty \int_0^x e^{-(x+y)} dy dx$

[Sugerencia: Sea $I_y(x) = \begin{cases} 1 & \text{si } y < x \\ 0 & \text{si } y \geq x \end{cases}$ utilice esta función para igualar esta integral a otra cuyos términos vayan de 0 a ∞ .]

10. Utilice la simulación para aproximar $\text{Cov}(U, e^U)$, donde U es uniforme en $(0, 1)$. Compare su aproximación con la respuesta exacta.
11. Sea U uniforme en $(0, 1)$. Utilice simulación para aproximar lo siguiente:
- (a) $\text{Cov}(U, \sqrt{1 - U^2})$.
- (b) $\text{Cov}(U^2, \sqrt{1 - U^2})$.
12. Para U_1, U_2, \dots variables aleatorias uniformes en $(0, 1)$, definimos

$$N = \text{Mínimo} \left\{ n: \sum_{i=1}^n U_i > 1 \right\}$$

Es decir, N es igual a la cantidad de números aleatorios que deben sumarse para exceder a 1.

- (a) Estime $E[N]$ generando 100 valores de N .
- (b) Estime $E[N]$ generando 1000 valores de N .
- (c) Estime $E[N]$ generando 10000 valores de N .
- (d) ¿Cuál cree que sea el valor de $E[N]$?

13. Sean $U_i, i \geq 1$, números aleatorios. Defina N como

$$N = \text{Máximo} \left\{ n: \prod_{i=1}^n U_i \geq e^{-3} \right\}$$

donde $\prod_{i=1}^0 U_i = 1$.

(a) Determine $E[N]$ por simulación.

(b) Determine $P\{N = i\}$, para $i = 0, 1, 2, 3, 4, 5, 6$, por simulación.

Bibliografía

Knuth, D., *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*, Reading, M. A., Addison-Wesley, 1981.

L'Ecuyer, P., "Random Numbers for Simulation", en *Commun. Assoc. Comput. Mach.*, **33**, 1990.

Marsaglia, G., "Random Numbers Fall Mainly in the Planes", en *Proc. Natl. Acad. Sci.*, **61**, 1962, pp. 25-28.

Marsaglia, G., "The Structure of Linear Congruential Sequences", en S. K. Zaremba (comp.), *Applications of Number Theory to Numerical Analysis*, Londres, Academic Press, 1972, pp. 249-255.

Naylor, T., *Computer Simulation Techniques*, Nueva York, Wiley, 1966.

Ripley, B., *Stochastic Simulation*, Nueva York, Wiley, 1986.

von Neumann, J., "Various Techniques Used in Connection with Random Digits, 'Monte Carlo Method'", en *U.S. National Bureau of Standards Applied Mathematics Series*, núm. 12, 1951, pp. 36-38.

CAPÍTULO

4

Generación de variables aleatorias discretas

4.1 El método de la transformada inversa

Suponga que queremos generar el valor de una variable aleatoria discreta X con función de masa de probabilidad

$$P\{X = x_j\} = p_j, \quad j = 0, 1, \dots, \sum_j p_j = 1$$

Para esto, generamos un número aleatorio U ; es decir, U está distribuido uniformemente en $(0, 1)$, y sea

$$X = \begin{cases} x_0 & \text{si } U < p_0 \\ x_1 & \text{si } p_0 \leq U < p_0 + p_1 \\ \vdots & \\ x_j & \text{si } \sum_{i=1}^{j-1} p_i \leq U < \sum_{i=1}^j p_i \\ \vdots & \end{cases}$$

Como $P\{a \leq U < b\} = b - a$ para $0 < a < b < 1$, tenemos que

$$P\{X = x_j\} = P\left\{\sum_{i=1}^{j-1} p_i \leq U < \sum_{i=1}^j p_i\right\} = p_j$$

y entonces X tiene la distribución deseada.

Observaciones

1. Podemos escribir lo anterior en forma algorítmica, como

Generar un número aleatorio U
 Si $U < p_0$ hacer $X = x_0$ y terminar
 Si $U < p_0 + p_1$ hacer $X = x_1$ y terminar
 Si $U < p_0 + p_1 + p_2$ hacer $X = x_2$ y terminar
 \vdots

2. Si los $x_i, i \geq 0$, están ordenados de modo que $x_0 < x_1 < x_2 < \dots$ y si F denota la función de distribución de X , entonces $F(x_k) = \sum_{i=0}^k p_i$ y entonces

$$X \text{ será igual a } x_j \quad \text{si} \quad F(x_{j-1}) \leq U < F(x_j)$$

En otras palabras, después de generar un número aleatorio U determinamos el valor de X hallando el intervalo $(F(x_{j-1}), F(x_j))$ en el que está U [o, de forma equivalente, hallando la inversa de $F(U)$]. Es por esta razón que el anterior se llama método de la transformada inversa discreta para generar X . ■

El tiempo necesario para generar una variable aleatoria discreta mediante este método es proporcional al número de intervalos en los que se realizan las búsquedas. Por esta razón, a veces vale la pena considerar los valores posibles x_j de X en orden decreciente de los p_j .

Ejemplo 4a Si queremos simular una variable aleatoria X tal que

$$p_1 = 0.20, \quad p_2 = 0.15, \quad p_3 = 0.25, \quad p_4 = 0.40 \quad \text{donde } p_j = P\{X = j\}$$

entonces podríamos generar U y hacer lo siguiente:

Si $U < 0.20$ hacemos $X = 1$ y terminamos
 Si $U < 0.35$ hacemos $X = 2$ y terminamos
 Si $U < 0.60$ hacemos $X = 3$ y terminamos
 En caso contrario, hacemos $X = 4$

Sin embargo, un procedimiento más eficaz es el siguiente:

Si $U < 0.40$ hacemos $X = 4$ y terminamos
 Si $U < 0.65$ hacemos $X = 3$ y terminamos
 Si $U < 0.85$ hacemos $X = 1$ y terminamos
 En caso contrario, hacemos $X = 2$ ■

Un caso en el que no es necesario buscar el intervalo adecuado en el cual se encuentra el número aleatorio es cuando la variable aleatoria deseada es la variable

aleatoria discreta uniforme. Es decir, suponga que queremos generar el valor de X entre los valores $1, \dots, n$, cada uno igualmente probable. Es decir, $P\{X = j\} = 1/n, j = 1, \dots, n$. El resultado anterior implica que podemos lograr esto al generar U y luego hacer

$$X = j \quad \text{si} \quad \frac{j-1}{n} \leq U < \frac{j}{n}$$

Por lo tanto, X será igual a j si $j-1 \leq nU < j$; o bien, en otras palabras,

$$X = \text{Ent}(nU) + 1$$

donde $\text{Ent}(x)$, algunas veces se denota como $[x]$, es la parte entera de x (es decir, el mayor entero menor o igual a x).

Las variables aleatorias discretas uniformes son importantes en la simulación, como lo indican los dos ejemplos siguientes.

Ejemplo 4b Generación de una permutación aleatoria. Suponga que estamos interesados en generar una permutación de los números $1, 2, \dots, n$ tal que todos los $n!$ posibles ordenamientos sean igualmente probables. Para ello, el siguiente algoritmo elige primero uno de los números $1, \dots, n$ al azar y luego lo coloca en la posición n ; enseguida, escoge al azar uno de los $n-1$ números restantes y lo sitúa en la posición $n-1$; entonces, elige al azar uno de los $n-2$ números restantes y lo coloca en la posición $n-2$, y así sucesivamente ("elegir un número al azar" significa que cada uno de los números restantes puede elegirse con la misma probabilidad). Sin embargo, para no tener que considerar con exactitud los números que faltan por colocar, es conveniente y eficiente mantenerlos en una lista ordenada y luego elegir al azar la posición del número y no el propio número. Es decir, comenzamos con cualquier ordenamiento inicial P_1, P_2, \dots, P_n y elegimos una de las posiciones $1, \dots, n$ al azar, para luego intercambiar el número en esa posición con el de la posición n . Ahora elegimos al azar una de las posiciones $1, \dots, n-1$ e intercambiamos el número en esta posición con el número en la posición $n-1$, y así sucesivamente.

Si recordamos que $\text{Ent}(kU) + 1$ toma alguno de los valores $1, 2, \dots, k$, cada uno con la misma probabilidad, vemos que el algoritmo anterior para generar una permutación aleatoria se puede escribir como sigue:

PASO 1: Sea P_1, P_2, \dots, P_n cualquier permutación de $1, 2, \dots, n$ (por ejemplo, podemos elegir $P_j = j, j = 1, \dots, n$).

PASO 2: Sea $k = n$.

PASO 3: Generar un número aleatorio U ; sea $I = \text{Ent}(kU) + 1$.

PASO 4: Intercambiar los valores de P_i y P_k .

PASO 5: Sea $k = k - 1$ y si $k > 1$, ir al paso 3.

PASO 6: P_1, \dots, P_n es la permutación aleatoria deseada.

Por ejemplo, supongamos que $n = 4$ y que la permutación inicial es 1, 2, 3, 4. Si el primer valor de I (que puede ser 1, 2, 3 o 4 con la misma probabilidad) es $I = 3$, entonces se intercambian los elementos en las posiciones 3 y 4, de modo que la nueva permutación es 1, 2, 4, 3. Si el siguiente valor de I es $I = 2$, entonces se intercambian los elementos de las posiciones 2 y 3, de modo que la nueva permutación es 1, 4, 2, 3. Si el valor final de I es $I = 2$, entonces la permutación final es 1, 4, 2, 3, y éste es el valor de la permutación arbitraria.

Una propiedad muy importante del algoritmo anterior es que también se puede utilizar para generar un subconjunto aleatorio, digamos de tamaño r , de los enteros 1, \dots , n . Basta seguir el algoritmo hasta llenar las posiciones $n, n-1, \dots, n-r+1$. Los elementos de estas posiciones son el subconjunto aleatorio (al hacerlo, siempre suponemos que $r \leq n/2$; ya que si $r > n/2$, entonces podríamos elegir un subconjunto aleatorio de tamaño $n-r$ y hacer que los elementos que no estén en ese conjunto sean el subconjunto aleatorio de tamaño r).

Debemos observar que la capacidad de generar un subconjunto aleatorio es de particular importancia en las pruebas médicas. Por ejemplo, supongamos que un centro médico planea una prueba con un nuevo medicamento diseñado para reducir el nivel de colesterol en la sangre. Para probar su eficacia, el centro médico ha reclutado 1000 voluntarios como sujetos de prueba. Para tomar en cuenta la posibilidad de que el nivel de colesterol en la sangre de los sujetos pudiera haber sido influido por factores externos (como el cambio en las condiciones climáticas), se ha decidido separarlos en dos grupos de tamaño 500: un grupo de *tratamiento* que recibe el medicamento y un grupo de *control* que recibe un placebo. Ni los voluntarios ni los que administran el medicamento sabrán quién está en cada grupo (tal prueba es *doble ciego*). Faltaría determinar cuáles de los voluntarios se deben elegir para formar el grupo de tratamiento. Es claro que uno desea que ambos grupos sean lo más similares posible en todos los aspectos, excepto porque los miembros del primer grupo recibirán el medicamento, mientras los del segundo grupo recibirán un placebo; pues en este caso sería posible concluir que cualquier diferencia de respuesta entre los grupos se debe realmente al medicamento. Hay un acuerdo general de que la mejor forma de lograrlo es elegir los 500 voluntarios del grupo de tratamiento de una manera completamente aleatoria. Es decir, la elección se debe realizar de modo que cada uno de los $\binom{1000}{500}$ subconjuntos de 500 voluntarios tenga la misma probabilidad de constituir el grupo de tratamiento. ■

Ejemplo 4c Cálculo de promedios. Suponga que queremos aproximar $\bar{a} = \sum_{i=1}^n a(i)/n$, donde n es grande y los valores $a(i)$, $i = 1, \dots, n$, son complicados y no tan fáciles de calcular. Una forma de hacerlo es observar que si X es una variable aleatoria discreta uniforme sobre los enteros 1, \dots , n , entonces la variable aleatoria $a(X)$ tiene una media dada por

$$E[a(X)] = \sum_{i=1}^n a(i)P\{X = i\} = \sum_{i=1}^n \frac{a(i)}{n} = \bar{a}$$

Por lo tanto, si para obtener k variables aleatorias discretas uniformes X_i , $i = 1, \dots, k$, generamos k números aleatorios U_i y hacemos $X_i = \text{Ent}(nU_i) + 1$, entonces cada una de las k variables aleatorias $a(X_i)$ tendrá media \bar{a} , lo cual, junto con la ley fuerte de los grandes números, implica que cuando k es grande (aunque mucho menor que n), el promedio de estos valores será aproximadamente igual \bar{a} . Por lo tanto, podemos aproximar \bar{a} , utilizando

$$\bar{a} \approx \sum_{i=1}^k \frac{a(X_i)}{k}$$

Otra variable aleatoria que se puede generar sin necesidad de buscar el intervalo donde caiga el número aleatorio es la geométrica.

Ejemplo 4d Recuerde que X es una variable aleatoria geométrica con parámetro p si

$$P\{X = i\} = pq^{i-1}, \quad i \geq 1, \text{ donde } q = 1 - p$$

Se puede pensar que X representa el tiempo del primer éxito, cuando se realizan ensayos independientes, de los que cada uno es un éxito con probabilidad p . Como

$$\begin{aligned} \sum_{i=1}^{j-1} P\{X = i\} &= 1 - P\{X > j-1\} \\ &= 1 - P\{\text{primeros } j-1 \text{ ensayos sean todos fracasos}\} \\ &= 1 - q^{j-1}, \quad j \geq 1 \end{aligned}$$

podemos generar el valor de X al generar un número aleatorio U y hacer X igual al valor j para el cual

$$1 - q^{j-1} \leq U < 1 - q^j$$

o, en forma equivalente, para el cual

$$q^j < 1 - U \leq q^{j-1}$$

Es decir, podemos definir X como

$$X = \text{Mín}\{j: q^j < 1 - U\}$$

Por lo tanto, con base en que el logaritmo es una función monótona, y entonces $0 < a < b$ es equivalente a $\log^*(a) < \log^*(b)$, tenemos que X se puede expresar como

$$\begin{aligned} X &= \text{Mín}\{j: j \log(q) < \log(1 - U)\} \\ &= \text{Mín}\left\{j: j > \frac{\log(1 - U)}{\log(q)}\right\} \end{aligned}$$

donde la última desigualdad cambia de signo, ya que $\log(q)$ es negativo para $0 < q < 1$. Por lo tanto, con la notación $\text{Ent}()$ podemos expresar X como

$$X = \text{Ent}\left(\frac{\log(1 - U)}{\log(q)}\right) + 1$$

Por último, al observar que $1 - U$ también está uniformemente distribuida en $(0, 1)$, tenemos que

$$X \equiv \text{Ent}\left(\frac{\log(U)}{\log(q)}\right) + 1$$

también es geométrica con parámetro p . ■

4.2 Generación de una variable aleatoria Poisson

La variable aleatoria X es Poisson con media λ si

$$p_i = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, \dots$$

La clave para aplicar el método de la transformada inversa para generar esta variable aleatoria es la siguiente identidad (demostrada en la sección 2.4 del capítulo 2):

$$p_{i+1} = \frac{\lambda}{i+1} p_i \quad i \geq 0 \quad (4.1)$$

Al aprovechar esta recursión para calcular las probabilidades Poisson conforme se necesiten, el algoritmo de la transformada inversa para generar una variable aleatoria

* $\log(x)$ representa el logaritmo natural de X . [N. del R.T.]

Poisson con media λ se puede expresar como sigue (la cantidad i se refiere al valor en cuestión; $p = p_i$ es la probabilidad de que X sea igual a i , y $F = F(i)$ es la probabilidad de que X sea menor o igual a i).

PASO 1: Generar un número aleatorio U .

PASO 2: $i = 0$, $p = e^{-\lambda}$, $F = p$.

PASO 3: Si $U < F$, hacer $X = i$ y terminar.

PASO 4: $p = \lambda p / (i + 1)$, $F = F + p$, $i = i + 1$.

PASO 5: Ir al paso 3.

(Debemos observar en lo anterior que cuando escribimos, por ejemplo, $i = i + 1$, no queremos decir que i sea igual a $i + 1$, sino que el valor de i se debe incrementar en 1.) Para ver que el algoritmo anterior realmente genera una variable aleatoria Poisson con media λ , observemos que primero genera un número aleatorio U y luego verifica si $U < e^{-\lambda} = p_0$. En tal caso, hace $X = 0$. En caso contrario, en el paso 4 calcula p_1 mediante la recursión (4.1). Ahora verifica si $U < p_0 + p_1$ (donde el lado derecho es el nuevo valor de F), y en tal caso hace $X = 1$, y así sucesivamente.

El algoritmo anterior verifica en forma sucesiva si el valor de Poisson es 0, luego si es 1, luego 2, etc. Así, el número de comparaciones necesarias será 1 más que el valor generado de la Poisson. Por lo tanto, en promedio, lo anterior necesitará realizar $1 + \lambda$ búsquedas. Aunque esto está bien si λ es pequeño, se puede mejorar en gran medida si λ es grande. De hecho, como una variable aleatoria Poisson con media λ tiene una gran probabilidad de tomar uno de los dos valores enteros más cercanos a λ , un algoritmo más eficiente verificaría primero uno de estos valores, en vez de partir de 0 y trabajar hacia arriba. El algoritmo 4-1, escrito en BASIC y que aparece en el Apéndice de programas, hace $I = \text{Ent}(\lambda)$ y calcula p_I (considera primero los logaritmos y luego eleva el resultado a la potencia e). Luego utiliza la ecuación (4.1) para determinar $F(I)$ en forma recursiva. Ahora, para generar una variable aleatoria Poisson con media λ genera un número aleatorio U , y observa si $X \leq I$ revisando si $U \leq F(I)$. Luego busca hacia abajo, a partir de $X = I$ en el caso en que $X \leq I$ y hacia arriba si $X = I + 1$ (el algoritmo 4-1 también evita la posible dificultad de que, debido a errores de redondeo, cuando λ es grande la computadora podría considerar $e^{-\lambda}$ igual a cero).

Observación El número de búsquedas necesarias para el algoritmo 4-1 es aproximadamente 1 más que la diferencia absoluta entre la variable aleatoria X y su media λ . Como para λ grande una Poisson es (por el teorema central del límite) aproximadamente normal con media y varianza iguales a λ , tenemos que

$$\begin{aligned}\text{Número promedio de búsquedas} &\approx 1 + E[|X - \lambda|] \quad \text{donde } X \sim N(\lambda, \lambda)^{\dagger} \\ &= 1 + \sqrt{\lambda} E\left[\frac{|X - \lambda|}{\sqrt{\lambda}}\right] \\ &\approx 1 + \sqrt{\lambda} E[|Z|] \quad \text{donde } Z \sim N(0, 1) \\ &= 1 + 0.798\sqrt{\lambda} \quad (\text{véase el ejercicio 12})\end{aligned}$$

Es decir, al utilizar el algoritmo 4-1, el número promedio de búsquedas crece con la raíz cuadrada de λ conforme λ crece cada vez más. ■

4.3 Generación de variables aleatorias binomiales

Suponga que queremos generar el valor de X , una variable aleatoria binomial (n, p) ; es decir, X tal que

$$P\{X = i\} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n$$

Para esto, nos valemos del método de la transformada inversa y de la identidad recursiva

$$P\{X = i + 1\} = \frac{n-i}{i+1} \frac{p}{1-p} P\{X = i\}$$

Si i denota el valor en cuestión, $pr = P\{X = i\}$ la probabilidad de que X sea igual a i , y $F = F(i)$ la probabilidad de que X sea menor o igual a i , el algoritmo se puede expresar de la manera siguiente:

Algoritmo de la transformada inversa para generar una variable aleatoria binomial (n, p)

- PASO 1: Generar un número aleatorio U .
 PASO 2: $c = p/(1-p)$, $i = 0$, $pr = (1-p)^n$, $F = pr$.
 PASO 3: Si $U < F$, hacer $X = i$ y terminar.
 PASO 4: $pr = [c(n-i)/(i+1)]pr$, $F = F + pr$, $i = i + 1$.
 PASO 5: Ir al paso 3.

[†]Empleamos la notación $X \sim F$ para indicar que X tiene la función de distribución F . El símbolo $N(\mu, \sigma^2)$ representa una distribución normal con media μ y varianza σ^2 .

El algoritmo anterior verifica primero si $X = 0$, luego si $X = 1$, y así sucesivamente. Por lo tanto, el número de búsquedas que realiza es 1 más que el valor de X . Por consiguiente, en promedio, se necesitan $1 + np$ búsquedas para generar X . Como una variable aleatoria binomial (n, p) representa el número de éxitos en n ensayos independientes, en los que cada uno es un éxito con probabilidad p , esto implica que esta variable aleatoria también se puede generar restando a n el valor de una variable aleatoria binomial $(n, 1-p)$ (¿por qué?). Así, cuando $p > 1/2$, podemos generar una variable aleatoria binomial $(n, 1-p)$ mediante el método anterior y restar su valor de n para obtener la generación deseada.

Observaciones

1. Otra forma de generar una variable aleatoria binomial (n, p) utiliza su interpretación como el número de éxitos en n ensayos independientes, al generar n números aleatorios U_1, \dots, U_n y luego hacer X igual al número de U_i que son menores o iguales a p . Si consideramos el i -ésimo ensayo como un éxito si $U_i < p$ y observamos que la probabilidad de este evento es igual a p , es fácil ver que esto produce una variable aleatoria binomial (n, p) . Sin embargo, este método requiere n números aleatorios y realiza n comparaciones, mientras que el algoritmo de la transformada inversa sólo necesita un número aleatorio y realiza, en promedio, $1 + np$ comparaciones (junto con igual número de divisiones).
2. Como en el caso Poisson, cuando la media np es grande es mejor determinar primero si el valor generado es menor o igual a $I \equiv \text{Ent}(np)$ o si es mayor que I . En el primer caso, uno debe comenzar la búsqueda con $X = I$, luego con $X = I - 1, \dots$, etc., mientras que en el segundo caso uno debe iniciar la búsqueda con $I + 1$ e ir hacia arriba. ■

4.4 La técnica de aceptación y rechazo

Suponga que tenemos un método eficiente para simular una variable aleatoria con función de masa de probabilidad $\{q_j, j \geq 0\}$. Podemos emplearlo como la base para simular a partir de la distribución que tiene función de masa $\{p_j, j \geq 0\}$ primero simulando una variable Y aleatoria con función de masa $\{q_j\}$ y luego aceptar este valor simulado con una probabilidad proporcional a p_j/q_j .

Específicamente, sea c una constante tal que

$$\frac{p_j}{q_j} \leq c \quad \text{para toda } j \text{ tal que } p_j > 0 \quad (4.2)$$

Ahora tenemos la siguiente técnica, llamada método de rechazo o método de aceptación y rechazo, para simular una variable aleatoria X con función de masa $p_j = P\{X = j\}$.

Método de rechazo

- PASO 1: Simular el valor de Y , con función de masa de probabilidad q_j .
 PASO 2: Generar un número aleatorio U .
 PASO 3: Si $U < p_j/cq_j$, hacer $X = Y$ y terminar. En caso contrario, regresar al paso 1.

El método de rechazo se representa en la figura 4.1.

Ahora demostramos que el método de rechazo funciona.

Teorema El algoritmo de aceptación y rechazo genera una variable aleatoria X tal que

$$P\{X = j\} = p_j, \quad j = 0, \dots$$

Además, el número de iteraciones del algoritmo necesarias para obtener X es una variable aleatoria geométrica con media c .

Demostración Para comenzar, determinaremos la probabilidad de que una única iteración produzca el valor aceptado j .

$$\begin{aligned} P\{Y = j, \text{ aceptada}\} &= P\{Y = j\}P\{\text{aceptada} \mid Y = j\} \\ &= q_j \frac{p_j}{cq_j} \\ &= \frac{p_j}{c} \end{aligned}$$

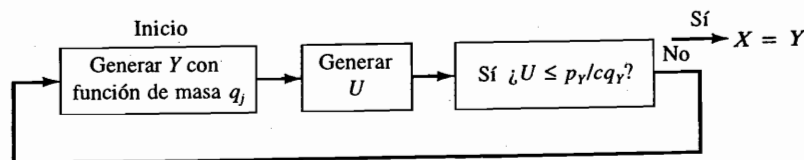


Figura 4.1 Aceptación y rechazo

Al sumar sobre j se obtiene la probabilidad de que una variable aleatoria generada sea aceptada:

$$P\{\text{aceptada}\} = \sum_j \frac{p_j}{c} = \frac{1}{c}$$

Como cada iteración produce de manera independiente un valor aceptado con probabilidad $1/c$, vemos que el número de iteraciones necesarias es una geométrica con media c . Además,

$$\begin{aligned} P\{X = j\} &= \sum_n P\{j \text{ aceptada en la iteración } n\} \\ &= \sum_n (1 - 1/c)^{n-1} \frac{p_j}{c} \\ &= p_j \end{aligned}$$

Observaciones

1. El lector debe observar que la forma en que “aceptamos el valor Y con probabilidad p_Y/cq_Y ” consiste en generar un número aleatorio U y luego aceptar Y si $U \leq p_Y/cq_Y$.
2. Como cada iteración del algoritmo produce de manera independiente un valor aceptado con probabilidad $P\{\text{aceptación}\} = K = 1/c$, esto implica que el número de iteraciones necesarias tendrá una distribución geométrica con parámetro $1/c$ y por tanto tendrá media c . Entonces, mientras más cerca esté c de 1 (nunca puede ser menor que 1, lo cual se puede ver al escribir la desigualdad (4.2) como $p_i \leq cq_i$ y sumar ambos lados) más eficaz será el método; esto es claro intuitivamente, pues mientras más cerca esté c de 1, más parecidas serán las dos funciones de masa $\{p_j\}$ y $\{q_j\}$. ■

Ejemplo 4e Suponga que queremos simular el valor de una variable aleatoria X que toma uno de los valores $1, 2, \dots, 10$ con probabilidades respectivas $0.11, 0.12, 0.09, 0.08, 0.12, 0.10, 0.09, 0.09, 0.10, 0.10$. Una posibilidad es utilizar el algoritmo de la transformada inversa, pero un planteamiento mejor consiste en usar el método de rechazo, con q como la densidad uniforme discreta en $1, \dots, 10$. Es decir, $q_j = 1/10$, $j = 1, \dots, 10$. Para esta elección de $\{q_j\}$, podemos elegir c como

$$c = \text{Máx} \frac{p_j}{q_j} = 1.2$$

de modo que el algoritmo sería el siguiente:

PASO 1: Generar un número aleatorio U_1 y hacer $Y = \text{Ent}(10U_1) + 1$.

PASO 2: Generar un segundo número aleatorio U_2 .

PASO 3: Si $U_2 \leq p_Y/0.12$, hacer $X = Y$ y terminar. En caso contrario, regresar al paso 1.

La constante 0.12 del paso 3 se debe a que $cq_Y = 1.2/10 = 0.12$. En promedio, este algoritmo requiere sólo 1.2 iteraciones para obtener el valor generado de X . ■

La fuerza del método de rechazo, una de cuyas primeras versiones fue propuesta por el famoso matemático John von Neumann, será aún más clara cuando consideremos su análogo al generar variables aleatorias continuas.

4.5 El método de composición

Suponga que tenemos un método eficiente para simular el valor de una variable aleatoria con una de las dos funciones de masa de probabilidad $\{p_j^1, j \geq 0\}$ o $\{p_j^2, j \geq 0\}$ y que queremos simular el valor de la variable aleatoria X con función de masa

$$P\{X = j\} = \alpha p_j^1 + (1 - \alpha)p_j^2, \quad j \geq 0 \quad (4.3)$$

donde $0 < \alpha < 1$. Una forma de simular esta variable aleatoria X es observar que si X_1 y X_2 son variables aleatorias con funciones de masa respectivas $\{p_j^1\}$ y $\{p_j^2\}$, entonces la variable aleatoria X definida como

$$X = \begin{cases} X_1 & \text{con probabilidad } \alpha \\ X_2 & \text{con probabilidad } 1 - \alpha \end{cases}$$

tendrá su función de masa dada por (4.3). Esto implica que para generar el valor de tal variable aleatoria primero generamos un número aleatorio U y luego un valor de X_1 si $U < \alpha$ y de X_2 si $U > \alpha$.

Ejemplo 4f Suponga que queremos generar el valor de una variable aleatoria X tal que

$$p_j = P\{X = j\} = \begin{cases} 0.05 & \text{para } j = 1, 2, 3, 4, 5 \\ 0.15 & \text{para } j = 6, 7, 8, 9, 10 \end{cases}$$

Si observamos que $p_j = 0.5p_j^1 + 0.5p_j^2$, donde

$$p_j^1 = 0.1, \quad j = 1, \dots, 10 \quad \text{y} \quad p_j^2 = \begin{cases} 0 & \text{para } j = 1, 2, 3, 4, 5 \\ 0.2 & \text{para } j = 6, 7, 8, 9, 10 \end{cases}$$

para ello primero generamos un número aleatorio U y después a partir de la uniforme discreta en $1, \dots, 10$ si $U < 0.5$ y de la uniforme discreta en $6, 7, 8, 9, 10$ en caso contrario. Es decir, podemos simular X como sigue:

PASO 1: Generar un número aleatorio U_1 .

PASO 2: Generar un número aleatorio U_2 .

PASO 3: Si $U_1 < 0.5$, sea $X = \text{Ent}(10U_2) + 1$. En caso contrario, sea $X = \text{Ent}(5U_2) + 6$. ■

Si $F_i, i = 1, \dots, n$ son funciones de distribución y $\alpha_i, i = 1, \dots, n$ son números no negativos cuya suma es 1, entonces la función de distribución F dada por

$$F(x) = \sum_{i=1}^n \alpha_i F_i(x)$$

es una *mezcla*, o *composición*, de las funciones de distribución $F_i, i = 1, \dots, n$. Una forma de simular a partir de F es simular una variable aleatoria I , igual a i con probabilidad $\alpha_i, i = 1, \dots, n$ y luego simular a partir de la distribución F_i (es decir, si el valor simulado de I es $I = j$, entonces la segunda simulación es a partir de F_j). Este método para simular a partir de F se conoce con frecuencia como el *método de composición*.

Ejercicios

1. Escriba un programa para generar n valores a partir de la función de masa de probabilidad $p_1 = 1/3, p_2 = 2/3$.
 - (a) Sea $n = 100$, ejecute el programa y determine la proporción de valores que sean iguales a 1.
 - (b) Repita (a) con $n = 1000$.
 - (c) Repita (a) con $n = 10000$.
2. Escriba un programa de computadora tal que, dada una función de masa de probabilidad $\{p_j, j = 1, \dots, n\}$ como entrada, proporcione como salida el valor de una variable aleatoria con esta función de masa.
3. Dé un algoritmo eficiente para simular el valor de una variable aleatoria X tal que

$$P\{X = 1\} = 0.3, \quad P\{X = 2\} = 0.2, \quad P\{X = 3\} = 0.35, \quad P\{X = 4\} = 0.15$$

4. Se baraja un conjunto de 100 cartas (numeradas del 1 al 100) y luego se voltean, una a la vez. Decimos que ocurre un "éxito" si la carta i es la i -ésima carta

volteada, $i = 1, \dots, 100$. Escriba un programa de simulación para estimar la esperanza y la varianza del número total de éxitos. Ejecute el programa. Determine las respuestas exactas y compárelas con sus estimaciones.

5. Otro método para generar una permutación aleatoria, diferente del presentado en el ejemplo 4b, consiste en generar de manera sucesiva una permutación aleatoria de los elementos $1, 2, \dots, n$ comenzando con $n = 1$, luego $n = 2$ y así sucesivamente (por supuesto, la permutación aleatoria cuando $n = 1$ es 1). Una vez que se tiene una permutación aleatoria de los primeros $n - 1$ elementos (digamos, P_1, \dots, P_{n-1}), para obtener la permutación aleatoria de los n elementos $1, \dots, n$ colocamos n en la posición final (para obtener la permutación P_1, \dots, P_{n-1}, n) e intercambiamos los elementos de la posición n (a saber, n) con el elemento de una posición elegida al azar; la posición 1, la posición 2, \dots , la posición n son igualmente probables.

- (a) Escriba un algoritmo que realice lo anterior.
 (b) Demuestre mediante inducción matemática sobre n que el algoritmo funciona, en el sentido de que la permutación obtenida tiene la misma probabilidad de ser cualquiera de las $n!$ permutaciones de $1, 2, \dots, n$.
6. (a) Emplee 100 números aleatorios para explicar la forma de determinar una aproximación de $\sum_{i=1}^N e^{i/N}$, donde $N = 10000$.
 (b) Obtenga la aproximación.
 (c) ¿Es buena su aproximación?
7. Se lanza de manera continua un par de dados legales, hasta que todos los posibles resultados $2, 3, \dots, 12$ hayan aparecido al menos una vez. Desarrolle un estudio de simulación para estimar el número esperado de lanzamientos necesarios.
8. Suponga que cada miembro en una lista de n elementos tiene un valor asociado, y sea $v(i)$ el valor asociado al i -ésimo elemento de la lista. Suponga que n es muy grande y también que cada elemento puede aparecer en lugares distintos de la lista. Explique la forma de utilizar números aleatorios para estimar la suma de los valores de los elementos distintos en la lista (en la que el valor de cada elemento se cuenta una sola vez, sin importar el número de veces que el elemento aparezca anotado).
9. Considere los eventos A_1, \dots, A_n donde los A_i , $i = 1, \dots, n$ consisten en los siguientes n_i resultados: $A_i = \{a_{i,1}, a_{i,2}, \dots, a_{i,n_i}\}$. Suponga que para cualquier resultado dado a , se conoce $P\{a\}$, la probabilidad de que el experimento produzca el resultado a . Explique la forma de utilizar los resultados del ejercicio 8 para estimar $P\{\cup_{i=1}^n A_i\}$, la probabilidad de que al menos uno de los eventos A_i ocurra. Observe que no suponemos que los eventos A_i , $i = 1, \dots, n$ sean mutuamente excluyentes.

10. La función de masa de probabilidad binomial negativa con parámetros (r, p) , donde r es un entero positivo y $0 < p < 1$, está dada por

$$p_j = \frac{(j-1)!}{(j-r)!(r-1)!} p^r (1-p)^{j-r}, \quad j = r, r+1, \dots$$

- (a) Por medio de la relación entre las variables aleatorias binomial negativa y geométrica y los resultados del ejemplo 4d obtenga un algoritmo de simulación de esta distribución.
 (b) Verifique la relación

$$p_{j+1} = \frac{j(1-p)}{j+1-r} p_j$$

- (c) Utilice la relación de la parte (b) para dar otro algoritmo que genere variables aleatorias binomiales negativas.
 (d) Utilice la interpretación de la distribución binomial negativa como el número de ensayos necesarios para acumular un total de r éxitos, cuando cada ensayo tiene éxito de manera independiente con probabilidad p , para obtener otro método para generar tal variable aleatoria.
11. Si Z es una variable aleatoria normal estándar, muestre que

$$E[|Z|] = \left(\frac{2}{\pi}\right)^{1/2} \approx 0.798$$

12. Dé dos métodos para generar una variable aleatoria X tal que

$$P\{X = i\} = \frac{e^{-\lambda} \lambda^i / i!}{\sum_{j=0}^k e^{-\lambda} \lambda^j / j!}, \quad i = 0, \dots, k$$

13. Sea X una variable aleatoria binomial con parámetros n y p . Suponga que queremos generar una variable aleatoria Y cuya función de masa de probabilidad es igual a la función de masa condicional de X dado que $X \geq k$ para alguna $k \leq n$. Sea $\alpha = P\{X \geq k\}$ y suponga que se ha calculado el valor de α .
 (a) Dé el método de transformada inversa para generar Y .
 (b) Dé otro método para generar Y .
 (c) ¿Para qué valores de α , pequeños o grandes, sería ineficiente el algoritmo en (b)?
14. Dé un método de simulación a partir de la función de masa de probabilidad p_j , $j = 5, 6, \dots, 14$, donde

$$p_j = \begin{cases} 0.11 & \text{cuando } j \text{ es impar y } 5 \leq j \leq 13 \\ 0.09 & \text{cuando } j \text{ es par y } 6 \leq j \leq 14 \end{cases}$$

15. Suponga que la variable aleatoria X puede tomar cualquiera de los valores $1, \dots, 10$ con probabilidades respectivas 0.06, 0.06, 0.06, 0.06, 0.06, 0.15, 0.13, 0.14, 0.15, 0.13. Utilice el método de composición para dar un algoritmo que genere el valor de X .
16. Presente un método para generar el valor de X , en el que

$$P\{X = j\} = \left(\frac{1}{2}\right)^{j+1} + \frac{\left(\frac{1}{2}\right)^{j-1}}{3^j}, \quad j = 1, 2, \dots$$

17. Sea X con función de masa $p_j = P\{X = j\}$, $\sum_{j=1}^{\infty} p_j = 1$. Sea

$$\lambda_n = P\{X = n | X > n-1\} = \frac{p_n}{1 - \sum_{j=1}^{n-1} p_j}, \quad n = 1, \dots$$

- (a) Muestre que $p_1 = \lambda_1$ y

$$p_n = (1 - \lambda_1)(1 - \lambda_2) \cdots (1 - \lambda_{n-1})\lambda_n$$

Las cantidades λ_n , $n \geq 1$, son las tasas discretas de riesgo, pues si pensamos en X como el tiempo de vida de algún artículo, entonces λ_n representa la probabilidad de que alguno que haya alcanzado la edad n muera durante ese periodo. El siguiente método para simular variables aleatorias discretas, llamado método de la tasa discreta de riesgo, genera una sucesión de números aleatorios y termina cuando el n -ésimo número aleatorio es menor que λ_n . El algoritmo se puede escribir como sigue:

PASO 1: $X = 1$.

PASO 2: Generar un número aleatorio U .

PASO 3: Si $U < \lambda_x$, terminar.

PASO 4: $X = X + 1$.

PASO 5: Ir al paso 2.

- (b) Muestre que el valor de X al terminar este proceso tiene la función de masa deseada.
- (c) Suponga que X es una variable aleatoria geométrica con parámetro p . Determine los valores λ_n , $n \geq 1$. Explique lo que hace el algoritmo anterior en este caso y por qué es evidente su validez.
18. Suponga que $0 \leq \lambda_n \leq \lambda$, para toda $n \geq 1$. Considere el siguiente algoritmo para generar una variable aleatoria con tasas discretas de riesgo $\{\lambda_n\}$.

PASO 1: $S = 0$.

PASO 2: Generar U y hacer $Y = \text{Ent}\left(\frac{\log(U)}{\log(1 - \lambda)}\right) + 1$.

PASO 3: $S = S + Y$.

PASO 4: Generar U .

PASO 5: Si $U \leq \lambda_S/\lambda$, hacer $X = S$ y terminar. En caso contrario, ir a 2.

(a) ¿Cuál es la distribución de Y en el paso 2?

(b) Explique lo que hace el algoritmo.

(c) Argumente que X es una variable aleatoria con tasas discretas de riesgo $\{\lambda_n\}$.

Generación de variables aleatorias continuas

Introducción

Todas las técnicas para generar una variable aleatoria discreta tienen su análogo en el caso continuo. En las secciones 5.1 y 5.2 presentamos el método de la transformada inversa y el método de rechazo para generar variables aleatorias continuas. En la sección 5.3 consideramos un método poderoso para generar variables aleatorias normales, conocido como el método polar. Por último, en las secciones 5.4 y 5.5 examinamos el problema de generar procesos Poisson y procesos Poisson no homogéneos.

5.1 El algoritmo de la transformada inversa

Consideremos una variable aleatoria continua con función de distribución F . Un método general para generarla, llamado de la transformación inversa, se basa en la siguiente proposición.

Proposición Sea U una variable aleatoria uniforme en $(0, 1)$. Para cualquier función de distribución continua F , invertible, la variable aleatoria X definida como

$$X = F^{-1}(U)$$

tiene distribución F . [F^{-1} se define como el valor de x tal que $F(x) = u$.]

Demostración Sea F_X la función de distribución de $X = F^{-1}(U)$. Entonces

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{F^{-1}(U) \leq x\} \end{aligned} \quad (5.1)$$

Ahora, como F es una función de distribución, se tiene que $F(x)$ es una función monótona creciente de x y por lo tanto la desigualdad " $a \leq b$ " es equivalente a la desigualdad " $F(a) \leq F(b)$ ". Así, de la ecuación (5.1), vemos que

$$\begin{aligned} F_X(x) &= P\{F^{-1}(U) \leq x\} \\ &= P\{U \leq F(x)\} \quad \text{pues } F(F^{-1}(U)) = U \\ &= F(x) \quad \text{pues } U \text{ es uniforme en } (0,1). \end{aligned}$$

La proposición anterior muestra entonces que para generar una variable aleatoria X a partir de la función de distribución continua F , generamos un número aleatorio U y hacemos entonces $X = F^{-1}(U)$.

Ejemplo 5a Suponga que queremos generar una variable aleatoria X con función de distribución

$$F(x) = x^n, \quad 0 < x < 1$$

Si hacemos $x = F^{-1}(u)$, entonces

$$u = F(x) = x^n, \text{ o, en forma equivalente, } x = u^{1/n}$$

Por lo tanto, para generar dicha variable aleatoria X generamos un número aleatorio U y luego hacemos $X = U^{1/n}$. ■

El método de la transformada inversa es una forma poderosa de generar variables aleatorias exponenciales, como muestra el siguiente ejemplo.

Ejemplo 5b Si X es una variable aleatoria exponencial con razón 1, entonces su función de distribución está dada por

$$F(x) = 1 - e^{-x}$$

Si hacemos $x = F^{-1}(u)$, entonces

$$u = F(x) = 1 - e^{-x}$$

o bien

$$1 - u = e^{-x}$$

o, al obtener los logaritmos,

$$x = -\log(1 - u)$$

Por lo tanto, para generar una exponencial con parámetro 1 generamos un número aleatorio U y hacemos

$$X = F^{-1}(U) = -\log(1 - U)$$

Ahorramos algo de tiempo si observamos que $1 - U$ también es uniforme en $(0, 1)$ y entonces $-\log(1 - U)$ tiene la misma distribución que $-\log U$. Es decir, el negativo del logaritmo de un número aleatorio se distribuye exponencialmente con razón 1.

Además, advierta que si X es exponencial con media 1, entonces, para cualquier c positiva, cX es exponencial con media c . Por lo tanto, una variable aleatoria exponencial X con razón λ (media $1/\lambda$) se obtiene al generar un número aleatorio U y hacer

$$X = -\frac{1}{\lambda} \log U$$

Observación Lo anterior también nos proporciona otro algoritmo para generar una variable aleatoria Poisson. Para comenzar, recordemos que un proceso Poisson con razón λ surge cuando los tiempos entre los eventos consecutivos son exponenciales independientes con razón λ (véase la sección 2.9 del capítulo 2). Para tal proceso $N(1)$, el número de eventos hasta el instante 1, se distribuye Poisson con media λ . Sin embargo, si $X_i, i = 1, \dots$, denota los tiempos sucesivos entre las llegadas, entonces el n -ésimo evento ocurrirá en el instante $\sum_{i=1}^n X_i$, de modo que el número de eventos hasta el instante 1 se puede expresar como

$$N(1) = \text{Máx}\left\{n: \sum_{i=1}^n X_i \leq 1\right\}$$

Es decir, el número de eventos hasta el instante 1 es igual a la máxima n para la cual el n -ésimo evento ha ocurrido antes del instante 1. (Por ejemplo, si el cuarto evento ocurrió antes del instante 1 pero el quinto no, entonces es evidente que habría un total de cuatro eventos para el instante 1.) Por lo tanto, con los resultados del ejemplo 5b, es posible, para generar $N = N(1)$, una variable aleatoria Poisson con media λ , producir números aleatorios U_1, \dots, U_n, \dots y hacer

$$\begin{aligned} N &= \text{Máx}\left\{n: \sum_{i=1}^n -\frac{1}{\lambda} \log U_i \leq 1\right\} \\ &= \text{Máx}\left\{n: \sum_{i=1}^n \log U_i \geq -\lambda\right\} \\ &= \text{Máx}\{n: \log(U_1 \cdots U_n) \geq -\lambda\} \\ &= \text{Máx}\{n: U_1 \cdots U_n \geq e^{-\lambda}\} \end{aligned}$$

Por lo tanto, una variable aleatoria Poisson N con media λ se puede obtener al generar de manera sucesiva números aleatorios hasta que su producto caiga debajo de $e^{-\lambda}$, y luego hacer N igual a 1 menos que la cantidad de números aleatorios necesarios. Es decir,

$$N = \text{Mín}\{n: U_1 \cdots U_n < e^{-\lambda}\} - 1$$

Los resultados del ejemplo 5b, junto con la relación entre las distribuciones gamma y exponencial, sirven para generar de manera eficiente una variable aleatoria gamma (n, λ) .

Ejemplo 5c Suponga que queremos generar el valor de una variable aleatoria gamma (n, λ) . Como la función de distribución F de tal variable aleatoria está dada por

$$F(x) = \int_0^x \frac{\lambda e^{-\lambda y} (\lambda y)^{n-1}}{(n-1)!} dy$$

no es posible dar una expresión con forma cerrada para su inversa. Sin embargo, podemos emplear el resultado de que una variable aleatoria gamma X con parámetros (n, λ) se puede considerar como la suma de n exponenciales independientes, cada una con razón λ (véase la sección 2.9 del capítulo 2), y utilizar el ejemplo 5b para generar X . Específicamente, podemos obtener una variable aleatoria gamma (n, λ) si generamos n números aleatorios U_1, \dots, U_n y hacemos entonces

$$\begin{aligned} X &= -\frac{1}{\lambda} \log U_1 - \cdots - \frac{1}{\lambda} \log U_n \\ &= -\frac{1}{\lambda} \log(U_1 \cdots U_n) \end{aligned}$$

donde el uso de la identidad $\sum_{i=1}^n \log x_i = \log(x_1 \cdots x_n)$ ahorra tiempo de cálculo, pues sólo hay que calcular un logaritmo y no n .

Los resultados del ejemplo 5c permiten obtener una forma eficiente para generar un conjunto de variables aleatorias exponenciales, generando primero su suma para luego, al establecer condiciones sobre esa suma, generar los valores individuales. Por ejemplo, podríamos generar X y Y , un par de exponenciales independientes e idénticamente distribuidas con media 1, generando primero $X + Y$ y luego utilizando el resultado (ejercicio 36 del capítulo 2) de que, dado que $X + Y = t$, la distribución condicional de X es uniforme en $(0, t)$. Entonces, se puede utilizar el siguiente algoritmo para generar un par de exponenciales con media 1.

PASO 1: Generar números aleatorios U_1 y U_2 .

PASO 2: Hacer $t = -\log(U_1 U_2)$.

PASO 3: Generar un número aleatorio U_3 .

PASO 4: $X = tU_3, Y = t - X$.

Al comparar esto con el método más directo de generar dos números aleatorios U_1 y U_2 y luego hacer $X = -\log U_1$ y $Y = -\log U_2$, se tiene que el algoritmo anterior ahorra

un cálculo logarítmico, con un costo de dos multiplicaciones y la generación de un número aleatorio.

También podemos generar k exponenciales independientes con media 1, al generar primero su suma, digamos con $-\log(U_1 \cdot \dots \cdot U_k)$, y luego generar $k - 1$ números aleatorios adicionales U_1, \dots, U_{k-1} , que deben ser ordenados. Si $U_{(1)} < U_{(2)} < \dots < U_{(k-1)}$ son sus valores ordenados, y si $-\log(U_1 \cdot \dots \cdot U_k) = t$, entonces las k exponenciales son

$$t[U_{(i)} - U_{(i-1)}], \quad i = 1, 2, \dots, k, \text{ donde } U_{(0)} \equiv 0, U_{(k)} \equiv t$$

5.2 El método de rechazo

Suponga que tenemos un método para generar una variable aleatoria con función de densidad $g(x)$. Podemos utilizarlo como base para generar a partir de la distribución continua con función de densidad $f(x)$: se genera Y a partir de g y luego se acepta este valor generado con una probabilidad proporcional a $f(Y)/g(Y)$.

En concreto, sea c una constante tal que

$$\frac{f(y)}{g(y)} \leq c \quad \text{para toda } y$$

Tenemos entonces la siguiente técnica (ilustrada en la figura 5.1) para generar una variable aleatoria con densidad f .

El método de rechazo

PASO 1: Generar Y con densidad g .

PASO 2: Generar un número aleatorio U .

PASO 3: Si, $U \leq \frac{f(Y)}{cg(Y)}$, hacer $X = Y$. En caso contrario, regresar al paso 1.

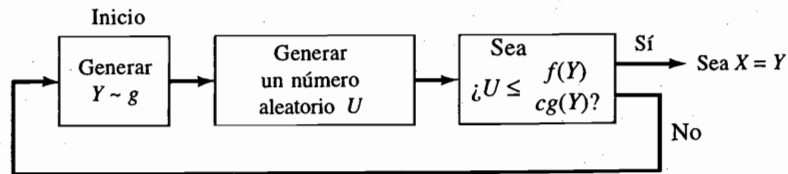


Figura 5.1 El método de rechazo para simular una variable aleatoria X con función de densidad f

El lector debe observar que el método de rechazo es exactamente igual al del caso de variables aleatorias discretas, con la única diferencia de que las densidades reemplazan a las funciones de masa. Tal como lo hicimos en el caso discreto, podemos demostrar el siguiente resultado.

Teorema

- (i) La variable aleatoria generada por el método de rechazo tiene densidad f .
- (ii) El número de iteraciones necesarias del algoritmo es una variable aleatoria geométrica con media c .

Como en el caso discreto, debe observarse que la forma en que para aceptar el valor Y con probabilidad $f(Y)/cg(Y)$ se genera un número aleatorio U y se acepta Y si $U \leq f(Y)/cg(Y)$.

Ejemplo 5d Emplearemos el método de rechazo para generar una variable aleatoria con función de densidad

$$f(x) = 20x(1 - x)^3, \quad 0 < x < 1$$

Como esta variable aleatoria (que es beta, con parámetros 2,4) se concentra en el intervalo $(0, 1)$, consideremos el método de rechazo con

$$g(x) = 1, \quad 0 < x < 1$$

Para determinar la constante c tal que $f(x)/g(x) \leq c$, utilizamos el cálculo para determinar el valor máximo de

$$\frac{f(x)}{g(x)} = 20x(1 - x)^3$$

Al derivar esta cantidad se obtiene

$$\frac{d}{dx} \left(\frac{f(x)}{g(x)} \right) = 20 [(1 - x)^3 - 3x(1 - x)^2]$$

Al igualar esto a cero, se tiene que el valor máximo se alcanza cuando $x = 1/4$ y entonces

$$\frac{f(x)}{g(x)} \leq 20 \left(\frac{1}{4} \right) \left(\frac{3}{4} \right)^3 = \frac{135}{64} \equiv c$$

Por lo tanto,

$$\frac{f(x)}{cg(x)} = \frac{256}{27} x(1 - x)^3$$

y así, el procedimiento de rechazo es el siguiente:

PASO 1: Generar números aleatorios U_1 y U_2 .

PASO 2: Si, $U_2 \leq \frac{256}{27} U_1 (1 - U_1)^3$, detenerse y hacer $X = U_1$. En caso contrario, rio, regresar al paso 1.

El número promedio de veces que se realiza el paso 1 es $c = \frac{135}{64} \approx 2.11$. ■

Ejemplo 5e Suponga que queremos generar una variable aleatoria con la densidad gamma $(\frac{3}{2}, 1)$

$$f(x) = Kx^{1/2}e^{-x}, \quad x > 0$$

donde $K = 1/\Gamma(\frac{3}{2}) = 2/\sqrt{\pi}$. Como esta variable aleatoria está concentrada en el eje positivo y tiene media $\frac{3}{2}$, es natural intentar la técnica de rechazo con una variable aleatoria exponencial con la misma media. Por lo tanto, sea

$$g(x) = \frac{2}{3}e^{-2x/3}, \quad x > 0$$

Ahora,

$$\frac{f(x)}{g(x)} = \frac{3K}{2}x^{1/2}e^{-x/3}$$

Al derivar e igualar a cero la derivada resultante, obtenemos que el valor máximo de este cociente se obtiene cuando

$$\frac{1}{2}x^{-1/2}e^{-x/3} = \frac{1}{3}x^{1/2}e^{-x/3}$$

es decir, cuando $x = \frac{3}{2}$. Por lo tanto,

$$\begin{aligned} c = \text{Máx} \frac{f(x)}{g(x)} &= \frac{3K}{2} \left(\frac{3}{2}\right)^{1/2} e^{-1/2} \\ &= \frac{3^{3/2}}{(2\pi e)^{1/2}} \quad \text{pues } K = 2/\sqrt{\pi} \end{aligned}$$

Como

$$\frac{f(x)}{cg(x)} = (2e/3)^{1/2} x^{1/2} e^{-x/3}$$

vemos que una variable aleatoria gamma $(\frac{3}{2}, 1)$ se puede generar como sigue:

PASO 1: Generar un número aleatorio U_1 y hacer $Y = -(\frac{3}{2})\log U_1$.

PASO 2: Generar un número aleatorio U_2 .

PASO 3: Si $U_2 < (2eY/3)^{1/2}e^{-Y/3}$, hacer $X = Y$. En caso contrario, regresar al paso 1.

El número promedio de iteraciones necesarias es

$$c = 3 \left(\frac{3}{2\pi e} \right)^{1/2} \approx 1.257. \quad \blacksquare$$

Observación Aunque la variable aleatoria gamma $(\frac{3}{2}, 1)$ tiene media $\frac{3}{2}$, no es claro de inmediato que deberíamos utilizar el rechazo con una exponencial que tenga la misma media. De hecho, suponga que hacemos

$$g(x) = \lambda e^{-\lambda x}$$

Entonces

$$\frac{f(x)}{g(x)} = \frac{Kx^{1/2}e^{-(1-\lambda)x}}{\lambda}$$

El valor máximo de este cociente se obtiene cuando

$$\frac{1}{2}x^{-1/2} = (1-\lambda)x^{1/2}$$

o bien, $x = [2(1-\lambda)]^{-1}$ [siempre que $\lambda < 1$; si $\lambda \geq 1$, es fácil ver que el cociente $f(x)/g(x)$ toma valores arbitrariamente grandes]. Por lo tanto, si utilizamos la exponencial con media $1/\lambda$, entonces el número promedio de iteraciones necesarias para el algoritmo es

$$c = \text{Máx} \frac{f(x)}{g(x)} = \frac{K}{\lambda} [2(1-\lambda)]^{-1/2} e^{-1/2}$$

Así, la mejor elección de λ es aquella que minimiza lo anterior o, en forma equivalente, que maximiza $\lambda(1-\lambda)^{1/2}$. El cálculo nos muestra de nuevo que este valor es tal que

$$(1-\lambda)^{1/2} = \frac{\lambda(1-\lambda)^{-1/2}}{2}$$

o, en forma equivalente,

$$1-\lambda = \lambda/2 \quad \text{o} \quad \lambda = \frac{2}{3}$$

Por lo tanto, la mejor exponencial que se puede utilizar en el método de rechazo para generar una variable aleatoria gamma $(\frac{3}{2}, 1)$ es la exponencial con media $\frac{3}{2}$. ■

Nuestro siguiente ejemplo muestra la forma de aplicar la técnica de rechazo para generar variables aleatorias normales.

Ejemplo 5f Generación de una variable aleatoria normal. Para generar una variable aleatoria normal estándar Z (es decir, aquella con media 0 y varianza 1), primero observe que el valor absoluto de Z tiene la función de densidad de probabilidad

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \quad 0 < x < \infty \quad (5.2)$$

Primero generamos a partir de la función de densidad anterior, valiéndonos del método de rechazo en el que g es la función de densidad exponencial con media 1, es decir,

$$g(x) = e^{-x} \quad 0 < x < \infty$$

Ahora,

$$\frac{f(x)}{g(x)} = \sqrt{2/\pi} e^{x-x^2/2}$$

y entonces el valor máximo de $f(x)/g(x)$ ocurre en el valor de x que maximiza $x - x^2/2$. El cálculo muestra que esto ocurre cuando $x = 1$, de modo que podemos considerar

$$c = \text{Máx} \frac{f(x)}{g(x)} = \frac{f(1)}{g(1)} = \sqrt{2e/\pi}$$

Como

$$\begin{aligned} \frac{f(x)}{cg(x)} &= \exp\left\{x - \frac{x^2}{2} - \frac{1}{2}\right\} \\ &= \exp\left\{-\frac{(x-1)^2}{2}\right\} \end{aligned}$$

esto implica que se obtiene el valor absoluto de una variable aleatoria normal unitaria como sigue:

PASO 1: Generar Y , una variable aleatoria exponencial con razón 1.

PASO 2: Generar un número aleatorio U .

PASO 3: Si $U \leq \exp\{-(Y-1)^2/2\}$, hacer $X = Y$. En caso contrario, regresar al paso 1.

Una vez simulada una variable aleatoria X con función de densidad como en la ecuación (5.2) (tal variable aleatoria se distribuye entonces como el valor absoluto de una normal unitaria), podemos obtener una normal unitaria Z al hacer que sea igualmente probable que Z tome X o $-X$.

En el paso 3, el valor Y se acepta si $U \leq \exp\{-(Y-1)^2/2\}$, lo cual es equivalente a $-\log U \geq (Y-1)^2/2$. Sin embargo, en el ejemplo 5b se mostró que $-\log U$ es exponencial con razón 1, de modo que lo anterior es equivalente a lo siguiente:

PASO 1: Generar exponenciales independientes con razón 1, Y_1 y Y_2 .

PASO 2: Si $Y_2 \geq (Y_1 - 1)^2/2$, hacer $X = Y_1$. En caso contrario, regresar al paso 1.

Suponga ahora que lo anterior implica que Y_1 es aceptada, por lo que sabemos que Y_2 es mayor que $(Y_1 - 1)^2/2$. ¿En cuánto excede una a la otra? Para responder, recordemos que Y_2 es exponencial con razón 1, por lo cual, dado que excede cierto valor, la cantidad en la que Y_2 excede a $(Y_1 - 1)^2/2$ [es decir, su "vida adicional" arriba del tiempo $(Y_1 - 1)^2/2$] se distribuye exponencial con razón 1, por la propiedad de que no tiene memoria. Es decir, cuando aceptamos en el paso 2 no sólo obtenemos X (el valor absoluto de una normal unitaria), sino que al calcular $Y_2 - (Y_1 - 1)^2/2$ también podemos generar una variable aleatoria exponencial (independiente de X) con razón 1.

Por lo tanto, en resumen, tenemos el siguiente algoritmo que genera una exponencial con razón 1 y una variable aleatoria normal estándar independiente.

PASO 1: Generar Y_1 , una variable aleatoria exponencial con razón 1.

PASO 2: Generar Y_2 , una variable aleatoria exponencial con razón 1.

PASO 3: Si $Y_2 - (Y_1 - 1)^2/2 > 0$, hacer $Y = Y_2 - (Y_1 - 1)^2/2$ e ir al paso 4. En caso contrario, ir al paso 1.

PASO 4: Generar un número aleatorio U y hacer

$$Z = \begin{cases} Y_1 & \text{si } U \leq \frac{1}{2} \\ -Y_1 & \text{si } U > \frac{1}{2} \end{cases}$$

Las variables aleatorias Z y Y generadas así son independientes, Z es normal con media 0 y varianza 1 y Y es exponencial con razón 1 (si desea que la variable aleatoria normal tenga media μ y varianza σ^2 , sólo considere $\mu + \sigma Z$). ■

Observaciones

1. Como, $c = \sqrt{2e/\pi} \approx 1.32$, lo anterior requiere un número de iteraciones del paso 2 con una distribución geométrica y media 1.32.
2. Si queremos generar una sucesión de variables aleatorias normales estándar, nos servimos de la variable aleatoria exponencial Y obtenida en el paso 3 como la exponencial inicial necesaria en el paso 1 para generar la siguiente normal. Por

lo tanto, en promedio, simulamos una normal estándar generando $1.64 (= 2 \times 1.32 - 1)$ exponenciales y calculando 1.32 cuadrados. ■

5.3 El método polar para generar variables aleatorias normales

Sean X y Y variables aleatorias normales unitarias independientes y sean R y θ las coordenadas polares del vector (X, Y) . Es decir (véase la figura 5.2)

$$R^2 = X^2 + Y^2$$

$$\tan \theta = \frac{Y}{X}$$

Como X y Y son independientes, su densidad conjunta es el producto de sus densidades individuales y por lo tanto está dada por

$$f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$= \frac{1}{2\pi} e^{-(x^2 + y^2)/2} \quad (5.3)$$

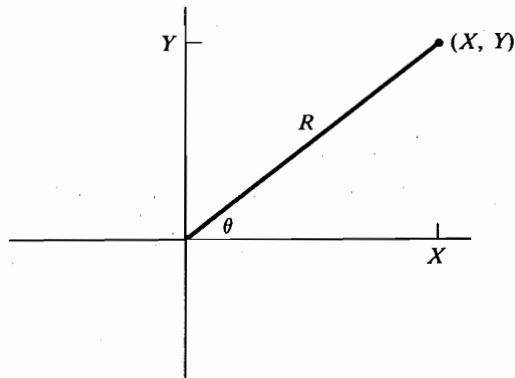


Figura 5.2

Para determinar la densidad conjunta de R^2 y θ —llamémosla $f(d, \theta)$ — hacemos el cambio de variables

$$d = x^2 + y^2, \quad \theta = \tan^{-1} \left(\frac{y}{x} \right)$$

Como es fácil ver que el jacobiano de esta transformación (es decir, el determinante de las derivadas parciales de d y θ con respecto de x y y) es igual a 2, la ecuación (5.3) implica que la función de densidad conjunta de R^2 y θ está dada por

$$f(d, \theta) = \frac{1}{2} \frac{1}{2\pi} e^{-d/2}, \quad 0 < d < \infty, \quad 0 < \theta < 2\pi$$

Sin embargo, como esto es igual al producto de una densidad exponencial con media 2 (a saber, $\frac{1}{2}e^{-d/2}$) y la densidad uniforme en $(0, 2\pi)$ (a saber, $1/2\pi$), esto implica que

$$R^2 \text{ y } \theta \text{ son independientes; } R^2 \text{ es exponencial con media 2} \quad (5.4)$$

$$\text{y } \theta \text{ se distribuye uniformemente en } (0, 2\pi)$$

Ahora podemos generar un par de variables aleatorias normales estándar independientes X y Y , utilizando (5.4) para generar primero sus coordenadas polares y luego transformarlas de nuevo en coordenadas rectangulares. Esto se logra de la manera siguiente:

PASO 1: Generar números aleatorios U_1 y U_2 .

PASO 2: $R^2 = -2 \log U_1$ (y entonces R^2 es exponencial con media 2). Sea $\theta = 2\pi U_2$ (y entonces θ es uniforme entre 0 y 2π).

PASO 3: Ahora, sean

$$X = R \cos \theta = \sqrt{-2 \log U_1} \cos(2\pi U_2) \quad (5.5)$$

$$Y = R \sin \theta = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

Las transformaciones dadas por las ecuaciones (5.5) se conocen como transformaciones de Box-Muller.

Por desgracia, el uso de las transformaciones de Box-Muller (5.5) para generar un par de normales unitarias independientes no es eficiente desde el punto de vista computacional, y la razón está en la necesidad de calcular las funciones trigonométricas seno y coseno. Sin embargo, hay una manera casual de evadir esta dificultad mediante un cálculo indirecto del seno y el coseno de un ángulo aleatorio (opuesto a un cálculo directo, el cual genera U y luego calcula el seno y el coseno de $2\pi U$). Para comenzar,

observe que si U es uniforme en $(0, 1)$, entonces $2U$ es uniforme en $(0, 2)$ y $2U - 1$ es uniforme en $(-1, 1)$. Así, si generamos números aleatorios U_1 y U_2 y hacemos

$$V_1 = 2U_1 - 1$$

$$V_2 = 2U_2 - 1$$

entonces (V_1, V_2) está uniformemente distribuida en el cuadrado de área 4 con centro en $(0, 0)$; véase la figura 5.3.

Ahora suponga que generamos de manera continua tales pares (V_1, V_2) hasta obtener uno que esté dentro del círculo de radio 1 con centro en $(0, 0)$; es decir, hasta que (V_1, V_2) es tal que $V_1^2 + V_2^2 \leq 1$. Ahora, se tiene que tal par (V_1, V_2) está uniformemente distribuido en el círculo. Si R y Θ denotan las coordenadas polares de este par, entonces no es difícil verificar que R y Θ son independientes y que R^2 está uniformemente distribuida en $(0, 1)$ (véase el ejercicio 19) y Θ está uniformemente distribuida en $(0, 2\pi)$. Como Θ es entonces un ángulo aleatorio, podemos generar el seno y el coseno de un ángulo aleatorio Θ generando un punto aleatorio (V_1, V_2) en el círculo y haciendo entonces

$$\sin \Theta = \frac{V_2}{R} = \frac{V_2}{(V_1^2 + V_2^2)^{1/2}}$$

$$\cos \Theta = \frac{V_1}{R} = \frac{V_1}{(V_1^2 + V_2^2)^{1/2}}$$

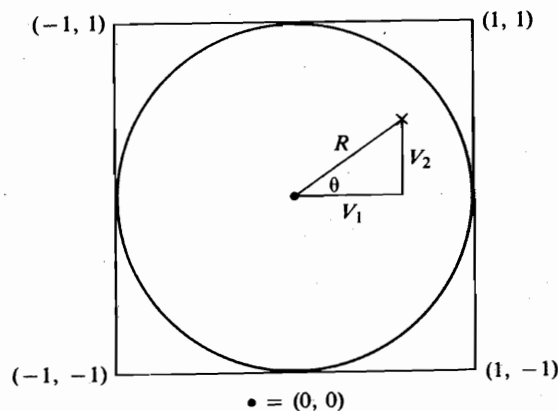


Figura 5.3

Ahora, gracias a la transformación de Box-Muller (5.5), para obtener normales unitarias independientes generamos un número aleatorio U y hacemos

$$X = (-2 \log U)^{1/2} \frac{V_1}{(V_1^2 + V_2^2)^{1/2}}$$

$$Y = (-2 \log U)^{1/2} \frac{V_2}{(V_1^2 + V_2^2)^{1/2}} \quad (5.6)$$

De hecho, como $R^2 = V_1^2 + V_2^2$ está uniformemente distribuida en $(0, 1)$ y es independiente del ángulo aleatorio Θ , podemos tomarla como el número aleatorio U necesario en las ecuaciones (5.6). Por lo tanto, si $S = R^2$, obtenemos que

$$X = (-2 \log S)^{1/2} \frac{V_1}{S^{1/2}} = V_1 \left(\frac{-2 \log S}{S} \right)^{1/2}$$

$$Y = (-2 \log S)^{1/2} \frac{V_2}{S^{1/2}} = V_2 \left(\frac{-2 \log S}{S} \right)^{1/2}$$

son normales unitarias independientes cuando (V_1, V_2) es un punto elegido al azar en el círculo de radio 1 con centro en el origen, y $S = V_1^2 + V_2^2$.

En resumen, tenemos el siguiente método para generar un par de normales unitarias independientes:

- PASO 1: Generar números aleatorios U_1 y U_2 .
- PASO 2: Hacer $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$.
- PASO 3: Si $S > 1$ regresar al paso 1.
- PASO 4: Regresar las normales unitarias independientes.

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2$$

El anterior se llama método polar. Como la probabilidad de que un punto aleatorio del cuadrado esté dentro del círculo es igual a $\pi/4$ (el área del círculo dividido entre el área del cuadrado), tenemos que, en promedio, el método polar necesitará $4/\pi = 1.273$ iteraciones del paso 1. Por lo tanto requerirá, en promedio, 2.546 números aleatorios, un logaritmo, una raíz cuadrada, una división y 4.546 multiplicaciones para generar dos normales unitarias independientes.

5.4 Generación de un proceso Poisson

Suponga que queremos generar los primeros n tiempos de evento de un proceso Poisson con razón λ . Para esto, nos servimos del resultado de que los tiempos entre los

eventos consecutivos de dicho proceso son variables aleatorias exponenciales independientes, cada una con razón λ . Así, una forma de generar el proceso es generar estos tiempos entre llegadas. Así, si generamos n números aleatorios U_1, U_2, \dots, U_n y $X_i = -1/\lambda \log U_i$, entonces X_i se puede considerar como el tiempo entre el $(i-1)$ -ésimo y el i -ésimo evento en el proceso Poisson. Como el tiempo real del j -ésimo evento será igual a la suma de los primeros j tiempos entre llegadas, los valores generados por los primeros n tiempos de evento son $\sum_{i=1}^j X_i, j = 1, \dots, n$.

Si quisiéramos generar las primeras T unidades de tiempo del proceso Poisson, podemos seguir el procedimiento anterior: generamos de manera sucesiva los tiempos entre las llegadas y nos detenemos cuando su suma excede a T . Es decir, con el siguiente algoritmo se generan todos los tiempos de eventos que ocurren en $(0, T)$ de un proceso Poisson con razón λ . En el algoritmo, t se refiere al tiempo, I es el número de eventos que han ocurrido hasta el instante t , y $S(I)$ es el tiempo del evento más reciente.

Generación de las primeras T unidades de tiempo de un proceso Poisson con razón λ

- PASO 1: $t = 0, I = 0$.
 PASO 2: Generar un número aleatorio U .
 PASO 3: $t = t - \frac{1}{\lambda} \log U$. Si $t < T$, terminar.
 PASO 4: $I = I + 1, S(I) = t$.
 PASO 5: Ir al paso 2.

El valor final de I en el algoritmo anterior representará el número de eventos que ocurren hasta el instante T , y los valores $S(1), \dots, S(I)$ serán los I tiempos de evento en orden creciente.

Hay otro método para simular las primeras T unidades de tiempo de un proceso Poisson que también es eficiente. Comienza simulando $N(T)$ (el número total de evento que ocurren hasta el tiempo T) y luego utiliza un resultado que establece que, dado $N(T)$, los tiempos en que ocurren estos eventos se distribuyen de manera independiente y uniforme en $(0, T)$ [esto es claro de manera intuitiva, pues por el axioma del incremento estacionario, es obvio que un tiempo de evento arbitrario está distribuido uniformemente en el intervalo, y por el axioma del incremento independiente es evidente que, dado $N(T)$, estos $N(T)$ tiempos de evento serían independientes]. Por lo tanto, podemos comenzar generando el valor de $N(T)$, una variable aleatoria Poisson con media λT (si λ es grande, por el método presentado en la sección 4.2 del capítulo 4; o bien, si λ es pequeña mediante ese método o el método bosquejado en la sección 5.2 de este capítulo). Si el valor generado de $N(T)$ es n , entonces generamos n números aleatorios, llamados U_1, \dots, U_n , y, como TU_i estarán uniformemente distribuidos

en $(0, T)$, el conjunto de tiempos de evento será entonces $\{TU_1, \dots, TU_n\}$. Si nos detuviéramos aquí, este método sería más eficiente que la simulación de los tiempos entre llegadas exponencialmente distribuidos. Sin embargo, por lo general queremos que los tiempos de evento estén en orden creciente [por ejemplo, para conocer $N(s)$ para toda $s < T$]; así, también necesitaríamos ordenar los valores $TU_i, i = 1, \dots, n$.

5.5 Generación de un proceso Poisson no homogéneo

Un proceso de conteo extremadamente importante para fines de modelación es el proceso Poisson no homogéneo, el cual relaja la hipótesis de incrementos estacionarios en el proceso Poisson. Con ello, permite que la tasa de llegada no sea constante, sino que varíe con el tiempo. Por lo general, es difícil obtener resultados analíticos para un modelo matemático que supone un proceso de llegada Poisson no homogéneo, y por ende tales procesos no se aplican con la frecuencia deseada. Sin embargo, como el uso de la simulación ayuda a analizar esos modelos, esperamos que tales modelos matemáticos sean cada vez más comunes.

Suponga que queremos simular las primeras T unidades de tiempo de un proceso Poisson no homogéneo con función de intensidad $\lambda(t)$. El primer método que presentamos, el método de *adelgazamiento* o *muestreo aleatorio*, comienza eligiendo un valor λ tal que

$$\lambda(t) \leq \lambda \quad \text{para toda } t \leq T$$

Ahora, como se mostró en el capítulo 2, tal proceso Poisson no homogéneo puede generarse mediante una selección aleatoria de los tiempos de evento de un proceso Poisson con razón λ . Es decir, si se cuenta un evento de un proceso Poisson con razón λ que ocurre en el instante t (de forma independiente a lo que ha ocurrido antes) con probabilidad $\lambda(t)/\lambda$, entonces el proceso de eventos contados es un proceso Poisson no homogéneo con función de intensidad $\lambda(t)$, $0 \leq t \leq T$. Por lo tanto, al simular un proceso Poisson y luego contar de manera aleatoria sus eventos, generamos el proceso Poisson no homogéneo. Esto se puede escribir como algoritmo de la manera siguiente.

Generación de las primeras T unidades de tiempo de un proceso Poisson no homogéneo

- PASO 1: $t = 0, I = 0$.
 PASO 2: Generar un número aleatorio U .
 PASO 3: $t = t - \frac{1}{\lambda} \log U$. Si $t > T$, terminar.
 PASO 4: Generar un número aleatorio U .

PASO 5: Si $U \leq \lambda(t)/\lambda_i$, hacer $I = I + 1$, $S(I) = t$.

PASO 6: Ir al paso 2.

Aquí, $\lambda(t)$ es la función de intensidad y λ es tal que $\lambda(t) \leq \lambda$. El valor final de I representa el número T de tiempos de evento, y $S(1), \dots, S(I)$ son los tiempos de evento.

Es claro que el procedimiento anterior, conocido como algoritmo de adelgazamiento (pues reduce los puntos Poisson homogéneos) es más eficiente, en el sentido de tener el menor número de tiempos de evento rechazados, cuando $\lambda(t)$ está cerca de λ en todo el intervalo. Así, una mejora obvia consiste en descomponer el intervalo en subintervalos y luego utilizar el procedimiento en cada subintervalo. Es decir, determinar valores adecuados k , $0 = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} = T$, $\lambda_1, \dots, \lambda_{k+1}$ tales que

$$\lambda(s) \leq \lambda_i, \quad \text{si } t_{i-1} \leq s < t_i, \quad i = 1, \dots, k+1 \quad (5.7)$$

Ahora, para generar el proceso Poisson no homogéneo sobre el intervalo (t_{i-1}, t_i) generamos variables aleatorias exponenciales con razón λ_i , y aceptamos el evento generado que ocurre en el instante s , $s \in (t_{i-1}, t_i)$, con probabilidad $\lambda(s)/\lambda_i$. Debido a la propiedad de la exponencial de no tener memoria y al hecho de que la razón de una exponencial se puede modificar multiplicando por una constante, no se pierde eficiencia al pasar de un subintervalo al siguiente. Es decir, si estamos en $t \in (t_{i-1}, t_i)$ y generamos X , una exponencial con razón λ_i , que es tal que $t + X > t_i$, entonces usamos $\lambda_i[X - (t_i - t)]/\lambda_{i+1}$ como la siguiente exponencial con razón λ_{i+1} .

Así, tenemos el siguiente algoritmo para generar las primeras T unidades de tiempo de un proceso Poisson no homogéneo con función de intensidad $\lambda(s)$ cuando se satisfacen las relaciones (5.7). En el algoritmo, t representa el tiempo actual, J el intervalo actual (es decir, $J = j$ cuando $t_{j-1} \leq t < t_j$), I es el número de eventos hasta el momento y $S(1), \dots, S(I)$ los tiempos de evento.

Generación de las primeras T unidades de tiempo de un proceso Poisson no homogéneo

PASO 1: $t = 0$, $J = 1$, $I = 0$.

PASO 2: Generar un número aleatorio U y hacer $X = \frac{-1}{\lambda_i} \log U$.

PASO 3: Si $t + X > t_j$, ir al paso 8.

PASO 4: $t = t + X$.

PASO 5: Generar un número aleatorio U .

PASO 6: Si $U \leq \lambda(t)/\lambda_j$, hacer $I = I + 1$, $S(I) = t$.

PASO 7: Ir al paso 2.

PASO 8: Si $J = k + 1$, terminar.

PASO 9: $X = (X - t_j + t)\lambda_j/\lambda_{j+1}$, $t = t_j$, $J = J + 1$.

PASO 10: Ir al paso 3.

Ahora supongamos que en algún subintervalo (t_{i-1}, t_i) se tiene que $\lambda_i > 0$, donde

$$\lambda_i \equiv \text{Ínfimo } \{\lambda(s): t_{i-1} \leq s < t_i\}$$

En tal caso, no debemos emplear el algoritmo de adelgazamiento en forma directa, sino primero simular un proceso Poisson con razón λ_i en el intervalo deseado y luego simular un proceso Poisson no homogéneo con la función de intensidad $\lambda(s) = \lambda(s) - \lambda_i$ cuando $s \in (t_{i-1}, t_i)$ (la última exponencial generada para el proceso Poisson, que va más allá de la frontera deseada, no tiene que desecharse, pues puede transformarse de manera adecuada para ser reutilizada). La superposición (o fusión) de los dos procesos da como resultado el proceso deseado en el intervalo. La razón para seguir este método es que ahorra la necesidad de generar variables aleatorias uniformes para un número distribuido en forma Poisson, con media $\lambda_i(t_i - t_{i-1})$, de los tiempos de evento. Por ejemplo, consideremos el caso donde

$$\lambda(s) = 10 + s, \quad 0 < s < 1$$

El uso del método de adelgazamiento con $\lambda = 11$ generaría un número esperado de 11 eventos, cada uno de los cuales necesitaría un número aleatorio para determinar si es aceptado o no. Por otro lado, generar un proceso Poisson con razón 10 y luego unirlo con un proceso Poisson no homogéneo con razón $\lambda(s) = s$, $0 < s < 1$ (generado por el algoritmo de adelgazamiento con $\lambda = 1$) produciría un número de tiempos de evento con la misma distribución, pero en el que el número esperado (que debería verificarse para ser aceptado) sería igual a 1.

El segundo método para simular un proceso Poisson no homogéneo con función de intensidad $\lambda(t)$, $t > 0$, consiste en generar de manera directa los tiempos de evento consecutivos. Así, sean S_1, S_2, \dots los tiempos de eventos consecutivos de tal proceso. Como es claro que estas variables aleatorias son dependientes, las generamos en serie, comenzando con S_1 , y luego tomamos el valor generado de S_1 para generar S_2 , y así sucesivamente.

Para comenzar, observamos que si un evento ocurre en el instante s , entonces, independientemente de lo ocurrido antes de s , el tiempo adicional hasta el siguiente evento tiene la distribución F_s , dada por

$$\begin{aligned}
 F_s(x) &= P\{\text{tiempo desde } s \text{ hasta que el siguiente evento es menor que } x \mid \text{evento en } s\} \\
 &= P\{\text{siguiente evento está antes de } x + s \mid \text{evento en } s\} \\
 &= P\{\text{evento entre } s \text{ y } s + x \mid \text{evento en } s\} \\
 &= P\{\text{evento entre } s \text{ y } s + x\} \text{ por incrementos independientes} \\
 &= 1 - P\{0 \text{ eventos en } (s, s + x)\} \\
 &= 1 - \exp\left(-\int_s^{s+x} \lambda(y) dy\right) \\
 &= 1 - \exp\left(-\int_0^x \lambda(s + y) dy\right) \quad (5.8)
 \end{aligned}$$

Ahora podemos simular los tiempos de evento S_1, S_2, \dots generando S_1 a partir de la distribución F_0 ; luego, si el valor simulado de S_1 es s_1 , generamos S_2 sumando s_1 a un valor generado a partir de la distribución F_{s_1} ; y si esta suma es s_2 , generamos S_3 sumando s_2 a un valor generado a partir de la distribución F_{s_2} ; y así sucesivamente. Por supuesto, el método utilizado para simular a partir de estas distribuciones dependería de su forma. En el siguiente ejemplo, es fácil invertir las distribuciones F_s de modo que se puede aplicar el método de la transformada inversa.

Ejemplo 5g Supongamos que $\lambda(t) = 1/(t + a)$, $t \geq 0$, para cierta constante positiva a . Entonces

$$\int_0^x \lambda(s + y) dy = \int_0^x \frac{1}{s + y + a} dy = \log\left(\frac{x + s + a}{s + a}\right)$$

Por lo tanto, de la ecuación (5.8),

$$F_s(x) = 1 - \frac{s + a}{x + s + a} = \frac{x}{x + s + a}$$

Para invertir esto, suponemos que $x = F_s^{-1}(u)$, y entonces

$$u = F_s(x) = \frac{x}{x + s + a}$$

o, en forma equivalente,

$$x = \frac{u(s + a)}{1 - u}$$

Ejercicios

Es decir,

$$F_s^{-1}(u) = (s + a) \frac{u}{1 - u}$$

Por lo tanto, para generar los tiempos de evento sucesivos S_1, S_2, \dots generamos números aleatorios U_1, U_2, \dots y definimos en forma recursiva

$$\begin{aligned}
 S_1 &= \frac{aU_1}{1 - U_1} \\
 S_2 &= S_1 + (S_1 + a) \frac{U_2}{1 - U_2} = \frac{S_1 + aU_2}{1 - U_2}
 \end{aligned}$$

y, en general,

$$S_j = S_{j-1} + (S_{j-1} + a) \frac{U_j}{1 - U_j} = \frac{S_{j-1} + aU_j}{1 - U_j}, \quad j \geq 2$$

Ejercicios

1. Dé un método para generar una variable aleatoria con función de densidad

$$f(x) = e^x/(e - 1), \quad 0 \leq x \leq 1$$

2. Dé un método para generar una variable aleatoria con función de densidad

$$f(x) = \begin{cases} \frac{x - 2}{2} & \text{si } 2 \leq x \leq 3 \\ \frac{2 - x/3}{2} & \text{si } 3 \leq x \leq 6 \end{cases}$$

3. Aplique el método de la transformada inversa para generar una variable aleatoria con función de distribución

$$F(x) = \frac{x^2 + x}{2}, \quad 0 \leq x \leq 1$$

4. Dé un método para generar una variable aleatoria con función de distribución

$$F(x) = 1 - \exp(-\alpha x^\beta), \quad 0 < x < \infty$$

Una variable aleatoria con tal distribución es una variable aleatoria Weibull.

5. Dé un método para generar una variable aleatoria con función de densidad

$$f(x) = \begin{cases} e^{2x}, & -\infty < x < 0 \\ e^{-2x}, & 0 < x < \infty \end{cases}$$

6. Sea X una variable aleatoria exponencial con media 1. Dé un algoritmo eficiente para simular una variable aleatoria cuya distribución es la distribución condicional de X , dado que $X < 0.05$. Es decir, su función de densidad es

$$f(x) = \frac{e^{-x}}{1 - e^{-0.05}}, \quad 0 < x < 0.05$$

Genere 1000 de tales variables y utilícelas para estimar $E[X | X < 0.05]$. Luego, determine el valor exacto de $E[X | X < 0.05]$.

7. (El método de composición) Suponga que es relativamente fácil generar variables aleatorias a partir de las distribuciones $F_i, i = 1, \dots, n$. ¿Cómo podríamos generar una variable aleatoria con la función de distribución

$$F(x) = \sum_{i=1}^n p_i F_i(x)$$

donde $p_i, i = 1, \dots, n$ son números no negativos cuya suma es 1?

8. Utilice los resultados del ejercicio 7 para dar algoritmos que generen variables aleatorias a partir de las siguientes distribuciones.

$$(a) F(x) = \frac{x + x^3 + x^5}{3}, \quad 0 \leq x \leq 1$$

$$(b) F(x) = \begin{cases} \frac{1 - e^{-2x} + 2x}{3} & \text{si } 0 < x < 1 \\ \frac{3 - e^{-2x}}{3} & \text{si } 1 < x < \infty \end{cases}$$

$$(c) F(x) = \sum_{i=1}^n \alpha_i x^i, \quad 0 \leq x \leq 1, \quad \text{donde } \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1$$

9. Dé un método para generar una variable aleatoria con función de distribución

$$F(x) = \int_0^{\infty} x^y e^{-y} dy, \quad 0 \leq x \leq 1$$

Sugerencia: Piense en términos del método de composición del ejercicio 7. En particular, sea F la función de distribución de X y suponga que la distribución condicional de X dado que $Y = y$ es

$$P\{X \leq x | Y = y\} = x^y, \quad 0 \leq x \leq 1$$

10. Escriba un algoritmo que se pueda utilizar para generar variables aleatorias exponenciales en conjuntos de 3. Compare los requisitos computacionales de este método con el presentado después del ejemplo 5c, que las genera en pares.

11. Suponga que es fácil generar una variable aleatoria a partir de cualquiera de las distribuciones $F_i, i = 1, \dots, n$. ¿Cómo podríamos generar a partir de las siguientes distribuciones?

$$(a) F(x) = \prod_{i=1}^n F_i(x)$$

$$(b) F(x) = 1 - \prod_{i=1}^n [1 - F_i(x)]$$

Sugerencia: Si $X_i, i = 1, \dots, n$ son variables aleatorias independientes, donde X_i tiene distribución F_i , ¿cuál variable aleatoria tiene como función de distribución a F ?

12. Utilice el método de rechazo y los resultados del ejercicio 11 para dar otros dos métodos, además del de la transformada inversa, que sirvan para generar una variable aleatoria con función de distribución

$$F(x) = x^n, \quad 0 \leq x \leq 1$$

Analice la eficiencia de los tres métodos para generar a partir de F .

13. Sea G una función de distribución con densidad g y suponga, para a y b constantes tales que $a < b$, que queremos generar una variable aleatoria a partir de la función de distribución

$$F(x) = \frac{G(x) - G(a)}{G(b) - G(a)}, \quad a \leq x \leq b$$

- (a) Si X tiene distribución G , entonces F es la distribución condicional de X dada ¿cuál información?

- (b) Muestre que el método de rechazo se reduce en este caso a generar una variable aleatoria X con distribución G y luego aceptarla si está entre a y b .

14. Dé dos métodos para generar una variable aleatoria con función de densidad

$$f(x) = xe^{-x}, \quad 0 \leq x < \infty$$

y compare su eficiencia.

15. Dé un algoritmo para generar una variable aleatoria con densidad

$$f(x) = 30(x^2 - 2x^3 + x^4), \quad 0 \leq x \leq 1$$

Analice la eficiencia de este método.

16. En el ejemplo 5f simulamos una variable aleatoria normal mediante la técnica de rechazo con una distribución exponencial de razón 1. Muestre que entre todas las funciones de densidad exponencial $g(x) = \lambda e^{-\lambda x}$, el número de iteraciones necesarias se minimiza cuando $\lambda = 1$.
17. Escriba un programa que genere variables aleatorias normales mediante el método del ejemplo 5f.
18. Escriba un programa que genere variables aleatorias normales mediante el método polar.
19. Sea (X, Y) uniformemente distribuida en un círculo de radio 1. Muestre que si R es la distancia del centro del círculo a (X, Y) , entonces R^2 es uniforme en $(0, 1)$.
20. Escriba un programa que genere las primeras T unidades de tiempo de un proceso Poisson con razón λ .
21. Para terminar su trabajo, un obrero debe pasar por k etapas en orden. El tiempo necesario para concluir la etapa i es una variable aleatoria exponencial con razón λ_i , $i = 1, \dots, k$. Sin embargo, después de concluir la etapa i , el obrero sólo pasará a la siguiente etapa con probabilidad α_i , $i = 1, \dots, k - 1$. Es decir, después de concluir la etapa i , el obrero dejará de trabajar con probabilidad $1 - \alpha_i$. Si X denota la cantidad de tiempo que dedica al trabajo, entonces X es una variable aleatoria de Cox. Escriba un algoritmo para generarla.
22. Los autobuses llegan a un encuentro deportivo de acuerdo con un proceso Poisson a razón de cinco por hora. Con la misma probabilidad, cada autobús puede transportar 20, 21, \dots , 40 aficionados y el número de autobuses distintos es independiente. Escriba un algoritmo para simular la llegada de aficionados al encuentro en el instante $t = 1$.
23. (a) Escriba un programa que utilice el algoritmo de adelgazamiento para generar las primeras 10 unidades de tiempo de un proceso Poisson no homogéneo con función de intensidad

$$\lambda(t) = 3 + \frac{4}{t+1}$$

- (b) Indique una forma de mejorar el algoritmo de adelgazamiento para este ejemplo.

24. Dé un algoritmo eficiente para generar las primeras 10 unidades de tiempo de un proceso Poisson no homogéneo con función de intensidad

$$\lambda(t) = \begin{cases} \frac{t}{5}, & 0 < t < 5 \\ 1 + 5(t - 5), & 5 < t < 10 \end{cases}$$

Bibliografía

- Dagpunar, T., *Principles of Random Variate Generation*, Oxford, Clarendon Press, 1988.
- Devroye, L., *Nonuniform Random Variate Generation*, Nueva York, Springer-Verlag, 1986.
- Fishman, G. S., *Principles of Discrete Event Simulation*, Nueva York, Wiley, 1978.
- Knuth, D., *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*, Reading, MA, Addison-Wesley, 1981.
- Law, A. M. y W. D. Kelton, *Simulation Modelling and Analysis*, Nueva York, McGraw-Hill, 2a. ed., 1991.
- Lewis, P. A. W. y G. S. Shedler, "Simulation of Nonhomogeneous Poisson Processes by Thinning", en *Nav. Res. Log. Quart.*, **26**, 1979, pp. 403-413.
- Marsaglia, G., "Generating Discrete Random Variables in a Computer", en *Commun. Assoc. Comput. Mach.*, **6**, 1963, pp. 37-38.
- Morgan, B. J. T., *Elements of Simulation*, Londres, Chapman and Hall, 1983.
- Ripley, B. D., "Computer Generation of Random Variables: A Tutorial", en *Inst. Statist. Rev.*, **51**, 1983, pp. 301-319.
- Ripley, B.D., *Stochastic Simulation*, Nueva York, Wiley, 1986.
- Rubenstein, R. Y., *Simulation and the Monte Carlo Method*, Nueva York, Wiley, 1981.
- Schmeiser, B. W., "Random Variate Generation, a Survey", en *Proc. 1980 Winter Simulation Conf.*, Orlando, FL, 1980, pp. 79-104.

El método de simulación por medio de eventos discretos

Introducción

La simulación de un modelo probabilístico requiere la generación de los mecanismos estocásticos del modelo para después observar el flujo resultante de éste conforme transcurre el tiempo. Según las razones para la simulación, habrá ciertas cantidades de interés que querramos determinar. Sin embargo, debido a que la evolución del modelo con el tiempo comprende a menudo una compleja estructura lógica de sus elementos, no siempre es fácil hallar la forma de seguirla como para determinar estas cantidades de interés. Así, para estos propósitos se ha elaborado un marco general, formulado en torno de la idea de “eventos discretos”. En tal marco se funda el *método de simulación con eventos discretos*.

6.1 Simulación mediante eventos discretos

Los elementos fundamentales en una simulación por medio de eventos discretos son las variables y los eventos. Para realizar la simulación, mantenemos un seguimiento continuo de ciertas variables. En general, hay tres clases de variables que se utilizan con frecuencia: la variable de tiempo, las variables de conteo o contadores y la variable de estado del sistema.

Variables

1. Variable de tiempo t

Se refiere al tiempo (simulado) que ha transcurrido

2. Variables de conteo

Estas variables mantienen un conteo del número de veces que ciertos eventos han ocurrido hasta el instante t

3. Variable de estado del sistema (ES)

Describe el “estado del sistema” en el instante t

Siempre que ocurre un “evento”, los valores de estas variables se modifican o actualizan, y como salida reunimos los datos de interés. Para determinar el momento de ocurrencia del siguiente evento, se mantiene una “lista de eventos” que enumera los próximos eventos y el momento en que se espera que ocurran. Siempre que “ocurra” un evento, reestablecemos la variable de tiempo, así como todas las variables de estado y contadores, y reunimos los datos importantes. De esta forma, podemos “seguir” el sistema conforme evoluciona con el tiempo.

Como lo anterior sólo pretende dar una idea general de los elementos de una simulación con eventos discretos, es útil analizar algunos ejemplos. En la sección 6.2 consideramos la simulación de un sistema de una línea de espera, o cola, con un único servidor. En las secciones 6.3 y 6.4 consideramos sistemas de colas con varios servidores. El modelo de la sección 6.3 supone que los servidores están ordenados en serie, y el de la sección 6.4 que están ordenados en paralelo. En la sección 6.5 consideramos un modelo de inventario y en 6.6 un problema de reparación de máquinas. En la sección 6.7 consideramos un modelo relativo a las opciones de acciones.

En todos los modelos de colas, suponemos que los clientes llegan de acuerdo con un proceso Poisson no homogéneo con una función de intensidad $\lambda(t)$ acotada, con $t > 0$. Al simular estos modelos, aplicamos la siguiente subrutina para generar el valor de una variable aleatoria T_s , definida como el tiempo de la primera llegada después del instante s .

Sea λ tal que $\lambda(t) \leq \lambda$ para toda t . Suponiendo dados $\lambda(t)$, $t > 0$ y λ , la siguiente subrutina genera el valor de T_s .

Una subrutina para generar T_s

PASO 1: Sea $t = s$.

PASO 2: Generar U .

PASO 3: Sea $t = t - \frac{1}{\lambda} \log U$.

PASO 4: Generar U .

PASO 5: Si $U \leq \lambda(t)/\lambda$, hacer $T_s = t$ y terminar.

PASO 6: Ir al paso 2.

6.2 Sistema de línea de espera con un servidor

Considere una estación de servicio a la cual los clientes llegan de acuerdo con un proceso Poisson no homogéneo con función de intensidad $\lambda(t)$, $\lambda \geq 0$. Hay un único servidor, y al llegar un cliente pasa a servicio si el servidor está libre en ese momento, o bien se une a la fila de espera si está ocupado. Cuando el servidor termina de dar servicio a un cliente, se ocupa del cliente que ha estado esperando más tiempo (la disciplina "primero en llegar, primero en atender") si hay clientes esperando, o bien, si no los hay, permanece libre hasta la llegada del siguiente cliente. El tiempo que tarda la atención a un cliente es una variable aleatoria (independiente de los demás tiempos de servicio y del proceso de llegada) con distribución de probabilidad G . Además, hay un tiempo fijo T después del cual no se permite que otras llegadas entren al sistema, aunque el servidor atiende a todos los que ya estén dentro del sistema en el instante T .

Suponga que estamos interesados en simular el sistema anterior para determinar cantidades tales como (a) el tiempo promedio que pasa un cliente dentro del sistema, y (b) el tiempo promedio posterior a T cuando sale el último cliente; es decir, el tiempo promedio en que el servidor puede ir a casa.

Para simular el sistema anterior utilizamos las siguientes variables:

1. Variable de tiempo t
2. Variables de conteo N_A : el número de llegadas (hasta el instante t)
 N_D : el número de salidas (hasta el instante t)
3. Variable de estado del sistema n : el número de clientes en el sistema (en el instante t)

Como el instante natural para modificar las cantidades anteriores es cuando ocurre una llegada o una salida, las consideramos "eventos"; es decir, hay dos tipos de eventos: llegadas y salidas. La lista de eventos contiene el instante de la siguiente llegada y el instante de la salida del cliente que se encuentra en servicio. Es decir, la lista de eventos es

$$LE = t_A, t_D$$

donde t_A es la hora de la siguiente llegada y t_D es la hora a la que concluye el servicio del cliente que se está atendiendo actualmente. Si no hay clientes en servicio, t_D es igual a ∞ .

Las variables de salida por registrar son $A(i)$, la hora de llegada del cliente i ; $D(i)$, la hora de salida del cliente i , y T_p , el tiempo de salida del último cliente, posterior a T .

Para comenzar la simulación, inicializamos las variables y los tiempos de los eventos como sigue:

Inicialización

Sean $t = N_A = N_D = 0$.

Sea $ES = 0$.

Generar T_0 , y hacer $t_A = T_0$, $t_D = \infty$.

Para actualizar el sistema, nos movemos a lo largo del eje del tiempo hasta encontrar el siguiente evento. Para ello, debemos considerar varios casos, dependiendo de cual sea el menor elemento de la lista de eventos. En lo sucesivo, Y se refiere a una variable aleatoria, con distribución G , para el tiempo de servicio.

$$t = \text{variable de tiempo}, \quad ES = n, \quad LE = t_A, t_D$$

Caso 1 $t_A \leq t_D$, $t_A \leq T$

Restablecer: $t = t_A$ (nos movemos hasta el tiempo t_A).

Restablecer: $N_A = N_A + 1$ (pues hay una llegada adicional en el instante t_A).

Restablecer: $n = n + 1$ (pues ahora se tiene un cliente más).

Generar T_p y hacer $t_A = T_p$ (ésta es la hora de la siguiente llegada).

Si $n = 1$, generar Y y hacer $t_D = t + Y$ (pues el sistema ha quedado vacío, por lo cual necesitamos generar el tiempo de servicio del nuevo cliente).

Reunir los datos de salida $A(N_A) = t$ (pues el cliente N_A llega en el instante t).

Caso 2 $t_D < t_A$, $t_D \leq T$

Restablecer: $t = t_D$.

Restablecer: $n = n - 1$.

Restablecer: $N_D = N_D + 1$ (pues ha ocurrido una salida en el instante t).

Si $n = 0$, hacer $t_D = \infty$; en caso contrario, generar Y y hacer $t_D = t + Y$.

Reunir los datos de salida $D(N_D) = t$ (pues el cliente N_D acaba de salir).

Caso 3 $\min(t_A, t_D) > T$, $n > 0$

Restablecer: $t = T$

Restablecer: $n = n - 1$

Restablecer: $N_D = N_D + 1$

Al llegar, el cliente pasará con el servidor 1 si éste se encuentra desocupado, o se formará en la fila de éste en caso contrario. De manera análoga, cuando el cliente termine el servicio en el servidor 1, entra a servicio con el servidor 2 si éste se encuentra libre, o se formará en su fila. Después de ser atendido en el servidor 2, el cliente sale del sistema. Los tiempos de servicio en el servidor i tienen la distribución G_i , $i = 1, 2$ (véase la figura 6.2).

Suponga que estamos interesados en realizar una simulación para estudiar la distribución de las cantidades de tiempo que un cliente pasa en ambos servidores. Para esto, utilizamos las siguientes variables.

Variable de tiempo t

Variable de estado del sistema (ES)

(n_1, n_2) : si hay n_1 clientes en el servidor 1 (incluyendo los formados y los que están en servicio) y n_2 clientes en el servidor 2

Variables de conteo

N_A : el número de llegadas hasta el instante t

N_D : el número de salidas hasta el instante t

Variables de salida

$A_1(n)$: la hora de llegada del cliente n , $n \geq 1$

$A_2(n)$: la hora de llegada del cliente n al servidor 2, $n \geq 1$

$D(n)$: la hora de salida del cliente n , $n \geq 1$

Lista de eventos t_A, t_1, t_2 , donde t_A es la hora de la siguiente llegada, y t_i es la hora a la que concluye el servicio del cliente actualmente atendido por el servidor i , $i = 1, 2$. Si

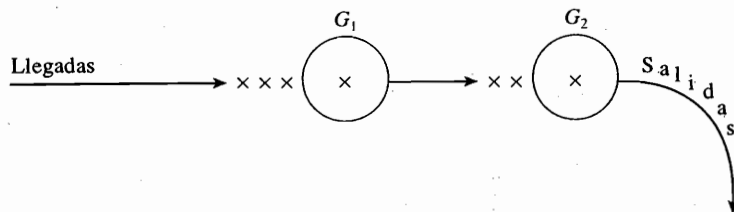


Figura 6.2

en un momento dado no se encuentra un cliente con el servidor i , entonces $t_i = \infty$, $i = 1, 2$. La lista de eventos siempre consta de las tres variables t_A, t_1, t_2 .

Para comenzar la simulación, inicializamos las variables y la lista de eventos como sigue:

Inicialización

Sean $t = N_A = N_D = 0$.

Sea $ES = (0, 0)$.

Generar T_0 y hacer $t_A = T_0$, $t_1 = t_2 = \infty$.

Para actualizar el sistema, nos movemos en el tiempo hasta encontrar el siguiente evento. Debemos considerar diversos casos, dependiendo del mínimo de los miembros de la lista de eventos. En lo sucesivo, Y_i se refiere a una variable aleatoria con distribución G_i , $i = 1, 2$.

$$ES = (n_1, n_2) \quad LE = t_A, t_1, t_2$$

Caso 1 $t_A = \min(t_A, t_1, t_2)$

Restablecer: $t = t_A$.

Restablecer: $N_A = N_A + 1$.

Restablecer: $n_1 = n_1 + 1$.

Generar T_n y hacer $t_A = T_n$.

Si $n_1 = 1$, generar Y_1 y hacer $t_1 = t + Y_1$.

Reunir los datos de salida $A_1(N_A) = t$.

Caso 2 $t_1 < t_A, t_1 \leq t_2$

Restablecer: $t = t_1$.

Restablecer: $n_1 = n_1 - 1$, $n_2 = n_2 + 1$.

Si $n_1 = 0$, hacer $t_1 = \infty$; en caso contrario, generar Y_1 y hacer $t_1 = t + Y_1$.

Si $n_2 = 1$, generar Y_2 y hacer $t_2 = t + Y_2$.

Reunir los datos de salida $A_2(N_A - n_1) = t$.

Caso 3 $t_2 < t_A, t_2 < t_1$

Restablecer: $t = t_2$.

Restablecer: $N_D = N_D + 1$.

Restablecer: $n_2 = n_2 - 1$.

Si $n_2 = 0$, hacer $t_2 = \infty$.

Si $n_2 > 0$, generar Y_2 , y hacer $t_2 = t + Y_2$.

Reunir los datos de salida $D(N_D) = t$.

Con este esquema de actualización, es fácil simular el sistema y reunir los datos de importancia.

6.4 Sistema de línea de espera con dos servidores en paralelo

Consideremos un modelo en el que los clientes llegan a un sistema con dos servidores. Al llegar, un cliente se forma en una fila si ambos servidores están ocupados, entra a servicio con el servidor 1 si ese servidor está desocupado, o bien con el servidor 2 en caso contrario. Cuando el cliente concluye el servicio con un servidor (sin importar cuál sea), sale del sistema y el cliente que ha estado formado más tiempo (si hay clientes en la cola) entra a servicio. La distribución de servicio en el servidor i es G_i , $i = 1, 2$ (véase la figura 6.3).

Suponga que queremos simular el modelo anterior manteniendo un registro de las cantidades de tiempo que pasa cada cliente dentro del sistema y el número de servicios realizados por cada servidor. Como hay varios servidores, los clientes no saldrán en el orden de llegada. Por lo tanto, para saber qué cliente deja el sistema al concluir su servicio, necesitamos llevar un registro para ver qué clientes están en el sistema. Así, los numeraremos conforme vayan llegando, de modo que el primero en llegar es el cliente número 1, el siguiente es el número 2, y así sucesivamente. Utilizaremos las siguientes variables:

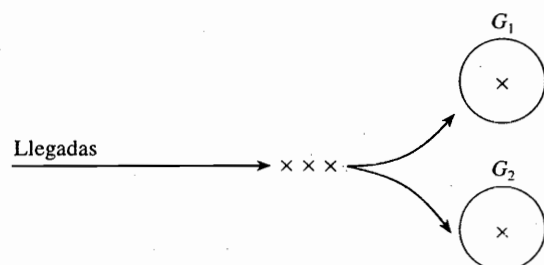


Figura 6.3

Variable de tiempo t

Variable de estado del sistema (ES)

$(n, i_1, i_2, \dots, i_n)$ si hay n clientes en el sistema, i_1 está con el servidor 1, i_2 está con el servidor 2, i_3 es el primero de la fila, i_4 es el siguiente, y así sucesivamente.

Observe que **ES** = (0) cuando el sistema está vacío, y **ES** = (1, j , 0) o (1, 0, j) cuando el único cliente es j y éste es atendido por el servidor 1 o el 2, respectivamente.

Variables de conteo

N_A : el número de llegadas hasta el instante t

C_j : el número de clientes atendidos por j , $j = 1, 2$, hasta el instante t

Variables de salida

$A(n)$: la hora de llegada del cliente n , $n \geq 1$

$D(n)$: la hora de salida del cliente n , $n \geq 1$

Lista de eventos t_A, t_1, t_2

Donde t_A es la hora de la siguiente llegada y t_i es la hora en la que concluye el servicio del cliente que en ese momento está siendo atendido por el servidor i , $i = 1, 2$. Si no hay un cliente actualmente con el servidor i , entonces hacemos $t_i = \infty$, $i = 1, 2$. En lo sucesivo, la lista de eventos siempre constará de las tres variables t_A, t_1, t_2 .

Para comenzar la simulación, inicializamos las variables y la lista de eventos como sigue:

Inicialización

Hacer $t = N_A = C_1 = C_2 = 0$.

Hacer **ES** = (0).

Generar T_0 y hacer $t_A = T_0$, $t_1 = t_2 = \infty$.

Para actualizar el sistema, nos movemos en el tiempo hasta encontrar el siguiente evento. En los siguientes casos, Y_i siempre se refiere a una variable aleatoria con distribución G_i , $i = 1, 2$.

Caso 1 $ES = (n, i_1, i_2, \dots, i_n)$ y $t_A = \min(t_A, t_1, t_2)$

Restablecer: $t = t_A$.

Restablecer: $N_A = N_A + 1$.

Generar T_i y hacer $t_A = T_i$.

Reunir los datos de salida $A(N_A) = t$.

Si $ES = (0)$:

Restablecer: $ES = (1, N_A, 0)$.

Generar Y_1 y hacer $t_1 = t + Y_1$.

Si $ES = (1, j, 0)$:

Restablecer: $ES = (2, j, N_A)$.

Generar Y_2 y hacer $t_2 = t + Y_2$.

Si $ES = (1, 0, j)$:

Restablecer: $ES = (2, N_A, j)$.

Generar Y_1 y hacer $t_1 = t + Y_1$.

Si $n > 1$:

Restablecer: $ES = (n + 1, i_1, \dots, i_n, N_A)$.

Caso 2 $ES = (n, i_1, i_2, \dots, i_n)$ y $t_1 < t_A, t_1 \leq t_2$

Restablecer: $t = t_1$.

Restablecer: $C_1 = C_1 + 1$.

Reunir los datos de salida $D(i_1) = t$.

Si $n_1 = 1$:

Restablecer: $ES = (0)$.

Restablecer: $t_1 = \infty$.

Si $n_1 = 2$:

Restablecer: $ES = (1, 0, i_2)$.

Restablecer: $t_1 = \infty$.

Si $n_1 > 2$:

Restablecer: $ES = (n - 1, i_3, i_2, \dots, i_n)$.

Generar Y_1 y hacer $t_1 = t + Y_1$.

Caso 3 $ES = (n, i_1, i_2, \dots, i_n)$ y $t_2 < t_A, t_2 < t_1$

Las actualizaciones en el caso 3 se dejan como ejercicio.

Si simulamos el sistema de acuerdo con lo anterior y detenemos la simulación en algún punto final determinado de antemano, entonces podemos utilizar las variables de salida, al igual que los valores finales de las variables de conteo C_1 y C_2 para obtener los datos relativos a las horas de llegada y salida de los diversos clientes, así como el número de servicios de cada servidor.

6.5 Modelo de inventario

Consideremos una tienda que almacena cierto producto, el cual vende a un precio unitario r . Los clientes que lo solicitan aparecen de acuerdo con un proceso Poisson con razón λ , y la cantidad que pide cada uno es una variable aleatoria con distribución G . Para cubrir la demanda, el propietario de la tienda debe tener a disposición una cantidad del producto y, siempre que el inventario disminuya, tendrá que ordenar más unidades al distribuidor. El propietario practica una política de solicitud (s, S) ; a saber, siempre que el inventario sea menor que s y no haya una solicitud previa, entonces pide determinada cantidad para que el inventario crezca hasta S , donde $s < S$. Es decir, si el nivel actual de inventario es x , no hay una solicitud pendiente y $x < s$, entonces se ordena la cantidad $S - x$. El costo de solicitud de y unidades del producto es una función dada $c(y)$, y se necesitan L unidades de tiempo para la entrega de un pedido; el pago se realiza al momento de la entrega. Además, la tienda paga un costo de mantenimiento del inventario de h por cada artículo, por unidad de tiempo. Suponga además que siempre que un cliente demande una cantidad mayor del producto de la existente, entonces se vende la cantidad a la mano y el resto del pedido representa una pérdida para la tienda.

Ahora veremos cómo aplicar la simulación para estimar la ganancia esperada de la tienda hasta cierto tiempo fijo T . Para esto, primero definimos las variables y los eventos de la manera siguiente.

Variable de tiempo t

Variable de estado del sistema (x, y)

donde x es la cantidad de inventario a la mano y y es la cantidad solicitada.

Variables de conteo

C , la cantidad total de costos de los pedidos hasta t

H , la cantidad total de costos de mantenimiento de inventario hasta t

R , la cantidad total de ingresos obtenidos hasta t

Los eventos serán la llegada de un cliente o de un pedido. Los tiempos de evento son

t_0 , el tiempo de llegada del siguiente cliente

t_1 , el tiempo de entrega de un pedido. Si no hay un pedido pendiente, entonces $t_1 = \infty$.

La actualización se realiza considerando el mínimo de los tiempos de evento. Si estamos en el instante t y tenemos los valores de las variables anteriores, entonces avanzamos en el tiempo de la manera siguiente.

Caso 1 $t_0 < t_1$

Restablecer: $H = H + (t_0 - t) x h$, pues entre los instantes t y t_0 incurrimos en un costo de mantenimiento de $(t_0 - t)h$ para cada una de las x unidades en el inventario.

Restablecer: $t = t_0$.

Generar D , una variable aleatoria con distribución G . D es la demanda del cliente que llegó en el instante t_0 .

Sea $w = \min(D, x)$ la cantidad del pedido por cubrir. El inventario posterior a esta solicitud es $x - w$.

Restablecer: $R = R + w r$.

Restablecer: $x = x - w$.

Si $x < s$ y $y = 0$, entonces $y = S - x$, $t_1 = t + L$.

Generar U y hacer $t_0 = t - \frac{1}{\lambda} \log(U)$.

Caso 2 $t_1 \leq t_0$

Restablecer: $H = H + (t_1 - t) x h$.

Restablecer: $t = t_1$.

Restablecer: $C = C + c(y)$.

Restablecer: $x = x + y$

Restablecer: $y = 0$, $t_1 = \infty$.

Este plan de actualización permite escribir un programa de simulación para analizar el modelo. Entonces, podríamos ejecutar la simulación hasta que ocurra el primer evento posterior a cierto valor del tiempo T predeterminado, grande, y luego emplear $(R - C - H)/T$ como estimación de la ganancia promedio de la tienda por unidad de tiempo. Al hacer esto para diversos valores de s y S , podríamos determinar una buena política de pedidos para la tienda.

6.6 Problema de reparación

Un sistema necesita que n máquinas estén funcionando. Para evitar las descomposturas, se dispone de algunas máquinas como repuestos. Siempre que una máquina se descompone, de inmediato se reemplaza por un repuesto y se envía al taller, en el cual una sola persona repara las máquinas una por una. Una vez reparada una máquina, queda disponible como repuesto para cuando surja la necesidad (véase la figura 6.4). Todos los tiempos de reparación son variables aleatorias independientes con una función de distribución común G . Cada vez que una máquina comienza a trabajar, el tiempo que funciona hasta descomponerse es una variable aleatoria, independiente de las anteriores, con función de distribución F .

El sistema "falla" cuando una máquina se descompone y no hay repuestos. Si al principio hay $n + s$ máquinas en buen estado, de las cuales n se ponen a trabajar y s

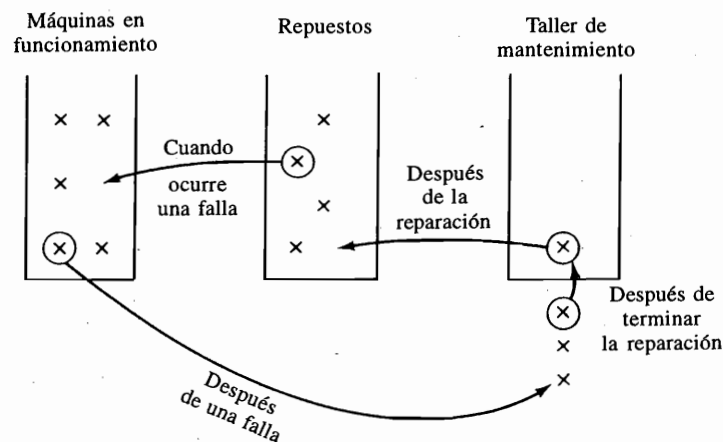


Figura 6.4 Modelo de reparación

quedan como repuestos, estamos interesados en simular el sistema para aproximar $E[T]$, donde T es el tiempo en que falla.

Para simular lo anterior, utilizamos las siguientes variables:

1. Variable de tiempo t
2. Variable de estado del sistema r : el número de máquinas descompuestas en el instante t

Como la variable de estado del sistema cambiará cuando una máquina en buen estado se descompone o cuando se concluye una reparación, decimos que ocurre un "evento" si ocurre cualquiera de estas situaciones. Para saber cuándo ocurrirá el siguiente evento, necesitamos llevar un registro de los instantes en los cuales fallan las máquinas que están en uso y el instante en que la máquina en reparación (si hay alguna) concluye su proceso. Puesto que siempre debemos determinar el mínimo de los n tiempos de descompostura, conviene anotar estos valores en una lista ordenada. Así, es conveniente que la lista de eventos sea como sigue:

Lista de eventos: $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_n, t^*$

donde t_1, \dots, t_n son los tiempos (en orden) de descompostura de las n máquinas y t^* es el tiempo en que la máquina en reparación vuelve a funcionar; si no hay una máquina en reparación en un momento dado, entonces $t^* = \infty$.

Para comenzar la simulación, inicializamos estas cantidades de la manera siguiente.

Inicialización

Sean $t = r = 0, t^* = \infty$.

Generar X_1, \dots, X_n , variables aleatorias independientes, cada una con distribución F .

Ordenar estos valores, de modo que t_i sea el i -ésimo menor, $i = 1, \dots, n$.

Sea la lista de eventos: t_1, \dots, t_n, t^* .

La actualización del sistema se realiza de acuerdo con los dos casos siguientes.

Caso 1 $t_1 < t^*$

Restablecer: $t = t_1$.

Restablecer: $r = r + 1$ (pues ha fallado otra máquina).

Si $r = s + 1$, detener esta ejecución y reunir los datos $T = t$ (pues como hay $s + 1$ máquinas descompuestas, no hay repuestos).

Si $r < s + 1$, generar una variable aleatoria X con distribución F . Esta variable aleatoria representará el tiempo de trabajo del repuesto que entrará en funciones.

Ahora reordenamos los valores $t_2, t_3, \dots, t_n, t + X$ y sea t_i el i -ésimo menor de ellos, $i = 1, \dots, n$.

Si $r = 1$, generar una variable aleatoria Y con función de distribución G y restablecer $t^* = t + Y$. (Esto es necesario debido a que en este caso, la máquina que acaba de fallar es la única descompuesta y por lo tanto de inmediato se comienza a reparar; Y será su tiempo de reparación y por lo tanto su reparación se concluirá en el instante $t + Y$.)

Caso 2 $t^* \leq t_1$

Restablecer: $t = t^*$.

Restablecer: $r = r - 1$.

Si $r > 0$, generar una variable aleatoria Y con función de distribución G , la cual representa el tiempo de reparación de la máquina que acaba de ingresar a servicio, y restablecer $t^* = t + Y$.

Si $r = 0$, hacer $t^* = \infty$.

Las reglas anteriores para la actualización aparecen en la figura 6.5.

Cada vez que nos detenemos (lo cual ocurre cuando $r = s + 1$), decimos que ha concluido una ejecución. La salida de la ejecución es el valor del tiempo de fallo T . Entonces, reinicializamos y simulamos otra ejecución. En resumen, realizamos un total, digamos, de k ejecuciones de modo que las variables de salida son, en forma sucesiva, T_1, \dots, T_k . Como estas k variables aleatorias son independientes y cada una representa un tiempo de falla, su promedio es decir, $\sum_{i=1}^k T_i / k$ es la estimación de $E[T]$, el tiempo medio de falla. La cuestión de determinar cuándo detener la simulación (es decir, determinar el valor de k) la analizamos en el capítulo 7, en el que se presenta los métodos para analizar de manera estadística la salida a partir de ejecuciones de simulación.

6.7 Ejercicio de opciones en acciones

Sea $S_n, n \geq 0$, el precio de una acción dada al final del día n . Un modelo común supone que

$$S_n = S_0 \exp\{X_1 + \dots + X_n\}, \quad n \geq 0$$

donde X_1, X_2, \dots es una sucesión de variables aleatorias normales e independientes, cada una con media μ y varianza σ^2 . El modelo supone que el incremento diario del precio con respecto del día anterior tiene una distribución común, llamada *modelo de cami-*

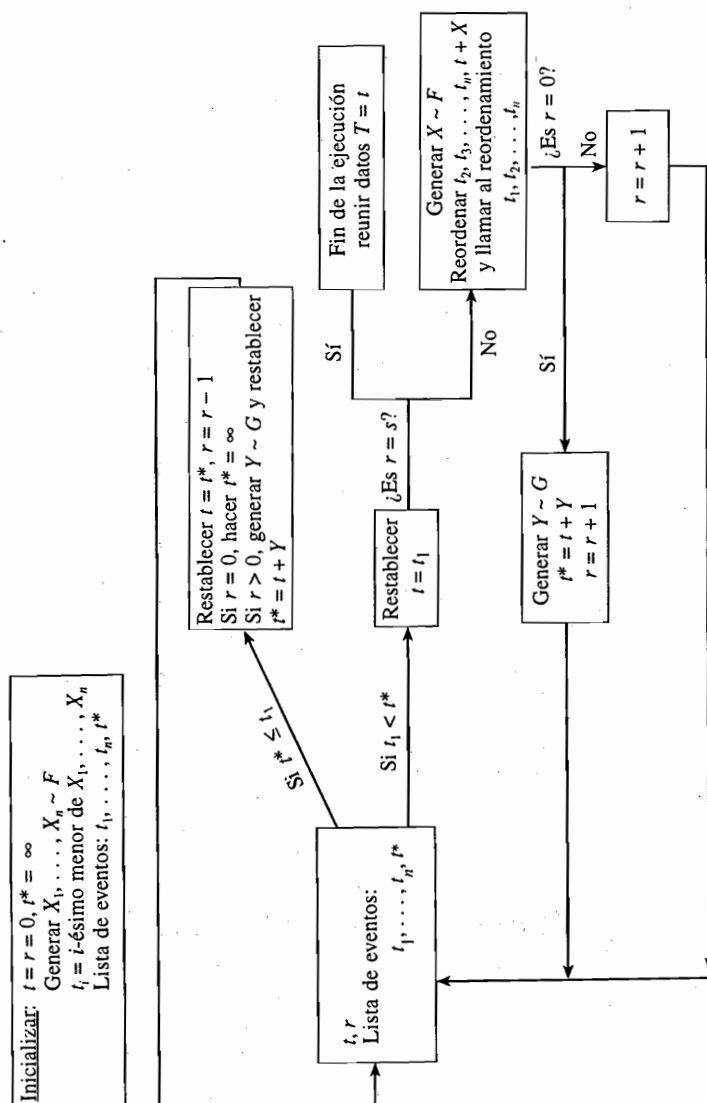


Figura 6.5

nata aleatoria lognormal. Sea $\alpha = \mu + \sigma^2/2$. Suponga ahora que usted tiene la opción de comprar una unidad de esta acción a un precio fijo K al final de cualquiera de los siguientes N días. Si usted ejerce esta opción cuando el precio de la acción es S , entonces, debido que usted sólo paga la cantidad K , llamamos a esto una ganancia de $S - K$ (pues teóricamente, usted podría cambiar de opinión de inmediato y vender la acción al precio S). La ganancia esperada al poseer la opción (la cual nunca debe ejercerse si el precio de la acción no excede a K durante el periodo en cuestión) depende de la política para ejercer opciones. Ahora, se puede mostrar que si $\alpha \geq 0$, entonces la política óptima consiste en esperar hasta el último momento posible y luego ejercer la opción si el precio es mayor que K y no ejercerla en caso contrario. Como $X_1 + \dots + X_N$ es una variable aleatoria normal con media $N\mu$ y varianza $N\sigma^2$, no es difícil calcular de manera explícita el resultado de esta política. Sin embargo, no es tan fácil caracterizar una política óptima, ni siquiera una que se acerque a la óptima, cuando $\alpha < 0$, y para cualquier política razonablemente buena, no se puede evaluar de manera explícita la ganancia esperada. Ahora daremos una política que se puede practicar cuando $\alpha < 0$ y que, aunque está lejos de ser óptima, parece razonablemente buena. Indica que la opción se debe ejercer cuando faltan m días por transcurrir siempre que, para cada $i = 1, \dots, m$, esa acción conduzca a un rédito esperado mayor que el obtenido al dejar que transcurran exactamente i días y luego ejercerla (si el precio en ese momento es mayor que K) o jamás ejercerla.

Sea $P_m = S_{N-m}$ el precio de la acción cuando faltan m días por transcurrir antes de que expire la opción. La política sugerida es la siguiente:

Política Si faltan m días por transcurrir, entonces ejercer la opción en este tiempo si

$$P_m > K$$

y, si para cada $i = 1, \dots, m$

$$P_m > K + P_m e^{i\alpha} \Phi(\sigma\sqrt{i} + b_i) - K\Phi(b_i)$$

donde

$$b_i = \frac{i\mu - \log(K/P_m)}{\sigma\sqrt{i}}$$

y donde $\Phi(x)$ es la función de distribución normal estándar, la cual puede aproximarse mediante la siguiente fórmula:

$$\Phi(x) \approx 1 - \frac{1}{\sqrt{2\pi}}(a_1y + a_2y^2 + a_3y^3)e^{1/2 - x^2/2}$$

donde

$$y = \frac{1}{1 + 0.33267x}$$

$$a_1 = 0.4361836$$

$$a_2 = -0.1201676$$

$$a_3 = 0.9372980$$

Sea SP el precio de la acción al ejercer la opción, si ésta se ejerce, y sea SP igual a K si la opción nunca se ejerce. Para determinar el valor esperado de esta política (es decir, determinar $E[SP] - K$), es necesario recurrir a la simulación. Para parámetros dados μ , σ , N , K , S_0 es suficiente simular el precio de la acción en días separados, generando X , una variable aleatoria con media μ y desviación estándar σ , y luego utilizar la relación

$$P_{m-1} = P_m e^X$$

Así, si P_m es el precio con m días faltantes y la política no pide ejercer la opción en este momento, entonces generaríamos X y determinaríamos el nuevo precio P_{m-1} , para que la computadora verifique si en este momento debe ejercerse la opción. En tal caso, para esa simulación se ejecuta $SP = P_{m-1}$; en caso contrario, determinaríamos el precio al final del siguiente día, y así sucesivamente. El valor promedio de $SP - K$, con respecto de un gran número de ejecuciones de simulación sería entonces nuestra estimación del valor esperado de contar con esta opción cuando se utiliza la política anterior.

6.8 Verificación del modelo de simulación

El producto final del método de eventos discretos para simulación es un programa de computadora que uno espera esté libre de errores. Por supuesto, para verificar que realmente no hay fallas en el programa, uno debe seguir todas las técnicas "estándares" de los programas depuradores. Sin embargo, hay varias técnicas que son particularmente aplicables en la depuración de los modelos de simulación, y ahora analizaremos algunas de ellas.

Como con todos los programas de gran tamaño, uno debe intentar depurar por "módulos" o subrutinas. Es decir, uno debe tratar de descomponer el programa en partes pequeñas y controlables, que sean cada una un todo lógico, para luego intentar depurar las partes. Por ejemplo, en los modelos de simulación, la generación de variables aleatorias constituye uno de estos módulos, cada uno de los cuales debe verificarse por separado.

La simulación siempre debe escribirse en forma amplia, con un gran número de variables de entrada. Con frecuencia, al elegir valores adecuados podemos reducir el modelo de

simulación a uno que pueda evaluarse de manera analítica o que ya haya sido estudiado con anterioridad, para comparar nuestros resultados simulados con respuestas conocidas.

En la etapa de verificación, el programa debe estar escrito para que el resultado contenga todas las cantidades aleatorias que genera. Al elegir de manera adecuada algunos valores particulares sencillos, podemos comparar la salida simulada con la respuesta obtenida a mano (y verificada por otros). Por ejemplo, suponga que estamos simulando las primeras T unidades de tiempo de un sistema de línea de espera con k servidores. Después de introducir los valores $T = 8$ (que pretende ser un número pequeño) y $k = 2$, suponga que el programa de simulación genera los siguientes datos:

Número de cliente:	1	2	3	4	5	6
Tiempo de llegada:	1.5	3.6	3.9	5.2	6.4	7.7
Tiempo de servicio:	3.4	2.2	5.1	2.4	3.3	6.2

y suponga que el programa da como resultado que el tiempo promedio ocupado por los seis clientes dentro del sistema es 5.12.

Sin embargo, al realizar los cálculos a mano (verificados por otros), vemos que el primer cliente ocupa 3.4 unidades de tiempo en el sistema; el segundo 2.2 (recuerde que hay dos servidores); el tercero llegó en el instante 3.9, entró a servicio en el instante 4.9 (cuando salió el primer cliente) y ocupó 5.1 unidades de tiempo en servicio; (así, el cliente 3 gastó un tiempo 6.1 en el sistema); el cliente 4 llegó al instante 5.2, entró a servicio en el instante 5.8 (cuando salió el número 2) y salió después de un tiempo adicional de 2.4; (así, el cliente 4 ocupó un tiempo 3.0 en el sistema); etc. A continuación presentamos estos cálculos:

Tiempo de llegada:	1.5	3.6	3.9	5.2	6.4	7.7
Tiempo de inicio del servicio:	1.5	3.6	4.9	5.8	8.2	10.0
Tiempo de salida:	4.9	5.8	10.0	8.2	11.5	16.2
Tiempo en el sistema:	3.4	2.2	6.1	3.0	5.1	8.5

Por lo tanto, la salida para el tiempo promedio ocupado dentro del sistema por todas las llegadas hasta el instante $T = 8$ sería

$$\frac{3.4 + 2.2 + 6.1 + 3.0 + 5.1 + 8.5}{6} = 4.71666 \dots$$

lo cual muestra que hay un error en el programa de computadora que dio el valor de salida 5.12.

Una técnica útil para buscar errores en el programa de computadora es utilizar un seguimiento o rastreo. En un rastreo, la variable de estado, la lista de eventos y las variables de conteo se imprimen después de la ocurrencia de cada evento. Esto permite seguir el sistema simulado con el tiempo para determinar si está funcionando como

se pretendía (si no aparecen errores evidentes, uno tendría que verificar los cálculos relativos a las variables de salida).

Ejercicios

1. Escriba un programa para generar la salida deseada para el modelo de la sección 6.2. Utilícelo para estimar el tiempo promedio que pasa un cliente en el sistema y el tiempo extra promedio del servidor, en el caso de que el proceso de llegada sea un proceso Poisson con razón 10, la densidad del tiempo de servicio es

$$g(x) = 20e^{-40x}(40x)^2, \quad x > 0$$

y $T = 9$. Intente primero con 100 ejecuciones y luego con 1000.

2. En el modelo de la sección 6.2, suponga que también queremos obtener información acerca del tiempo de inactividad diaria del servidor. Explique cómo hacerlo.
3. Suponga que los trabajos llegan a un sistema de línea de espera con un servidor de acuerdo con un proceso Poisson no homogéneo, cuya razón inicial es 4 por hora y que crece con una velocidad constante hasta llegar a 19 por hora después de 5 horas, para luego descender en proporción constante hasta 4 por hora, después de otras 5 horas. La razón comienza a repetirse de esta manera; es decir, $\lambda(t + 10) = \lambda(t)$. Suponga que la distribución del servicio es exponencial con razón de 25 por hora. Suponga además que siempre que el servidor concluya un servicio y no encuentre trabajos esperando, entra en un receso que se distribuye de manera uniforme en $(0, 0.3)$. Si al regresar de este receso no hay trabajos esperando, entonces toma otro receso. Utilice la simulación para estimar la cantidad esperada de tiempo que el servidor estará en receso durante las primeras 100 horas de operación. Realice 500 ejecuciones de simulación.
4. Complete el esquema de actualización para el caso 3 en el modelo de la sección 6.4.
5. Considere un modelo de línea de espera con un servidor, en el cual los clientes llegan de acuerdo con un proceso Poisson no homogéneo. Al llegar, entran a servicio si el servidor está desocupado o bien se forman en una fila. Sin embargo, suponga que cada cliente sólo puede permanecer formado una cantidad aleatoria de tiempo, con una distribución F , antes de salir del sistema. Sea G la distribución del servicio. Defina las variables y los eventos para analizar este modelo y dé los procedimientos de actualización. Suponga que estamos interesados en estimar el número promedio de clientes perdidos hasta el instante T ; un cliente se considera perdido si se va antes de recibir servicio.

6. En el ejercicio 5, suponga que el proceso de llegada es un proceso Poisson con razón 5; F es la distribución uniforme en $(0, 5)$ y G es una variable aleatoria exponencial con razón 4. Realice 500 ejecuciones de simulación para estimar el número esperado de clientes perdidos hasta el instante 100. Suponga que los clientes reciben servicio según su orden de llegada.
7. Repita el ejercicio 6, pero suponiendo que cada vez que el servidor concluye un servicio, el siguiente cliente por atender es aquél que tiene el menor tiempo de salida de la fila. Es decir, si dos clientes están esperando y uno saldrá de la fila si su servicio no ha comenzado antes del instante t_1 y el otro saldrá si el servicio no comienza antes del instante t_2 , entonces el primero entra a servicio si $t_1 < t_2$ y el segundo entra a servicio en caso contrario. ¿Piensa que esto incrementará o disminuirá el número promedio de clientes que salen sin recibir servicio?
8. En el modelo de la sección 6.4, suponga que G_1 es la distribución exponencial con razón 4 y que G_2 es exponencial con razón 3. Suponga que las llegadas constituyen un proceso Poisson con razón 6. Escriba un programa de simulación para generar datos correspondientes a las primeras 1000 llegadas. Empléelo para estimar:
 - (a) El tiempo promedio de estos clientes dentro del sistema.
 - (b) La proporción de servicios realizados por el servidor 1.
 - (c) Realice otra simulación de las primeras 1000 llegadas y utilícela para responder las partes (a) y (b). Compare sus respuestas con las ya obtenidas.
9. En el modelo de dos servidores en paralelo de la sección 6.4, suponga que cada servidor tiene su propia cola y que al llegar un cliente se forma en la menor. Si al llegar un cliente ambas colas tienen el mismo tamaño (o ambas son vacías), va al servidor 1.
 - (a) Determine las variables y los eventos adecuados para analizar este modelo y dé el procedimiento de actualización.

Con las mismas distribuciones y parámetros del ejercicio 8, determine:

 - (b) El tiempo promedio dentro del sistema para los primeros 1000 clientes.
 - (c) La proporción de los primeros 1000 servicios realizados por el servidor 1.

Antes de ejecutar su programa, ¿espera que sus respuestas de las partes (b) y (c) sean mayores o menores que las respuestas correspondientes del ejercicio 8?
10. En el ejercicio 9, suponga que cada llegada se envía al servidor 1 con probabilidad p , independiente de todo lo demás.
 - (a) Determine las variables y los eventos adecuados para analizar este modelo y dé el procedimiento de actualización.
 - (b) Utilice los parámetros del ejercicio 9, considere p igual a su estimación de la parte (c) de dicho ejercicio y simule el sistema para estimar las cantidades

definidas en las partes (b) y (c) del ejercicio 9. ¿Espera que sus respuestas a estos problemas sean mayores o menores que las obtenidas en ese ejercicio?

11. Para el modelo de reparación de la sección 6.6:

- (a) Escriba un programa de computadora para este modelo.
- (b) Ejecútelo para estimar el tiempo medio de falla en caso de que
 $n = 4, s = 3, F(x) = 1 - e^{-x}, \text{ y } G(x) = 1 - e^{-2x}.$

12. En el modelo de la sección 6.6, suponga que el taller tiene dos servidores, de los que cada uno ocupa un tiempo aleatorio con distribución G para dar servicio a una máquina descompuesta. Trace un diagrama de flujo para este sistema.

13. Un sistema experimenta un choque que ocurre de acuerdo con un proceso Poisson, a razón de uno por hora. Cada choque causa cierto daño. Suponemos que estos daños son variables aleatorias independientes (que además son independientes de los instantes en que ocurren), con la función de densidad común

$$f(x) = xe^{-x}, \quad x > 0$$

Los daños se disipan con el tiempo a una razón exponencial α ; es decir, un choque cuyo daño inicial es x tendrá un valor de daño restante igual a $xe^{-\alpha s}$ en el instante s después de haber ocurrido. Además, los valores de daño son acumulativos (así, por ejemplo, si hasta el instante t ha habido un total de dos choques, originados en los instantes t_1 y t_2 con daños iniciales x_1 y x_2 , entonces el daño total en el instante t es $\sum_{i=1}^2 x_i e^{-\alpha(t-t_i)}$). El sistema falla cuando el daño total excede cierta constante fija C .

- (a) Suponga que estamos interesados en realizar un estudio de simulación para estimar el tiempo promedio de falla del sistema. Defina los "eventos" y las "variables" del modelo y trace un diagrama de flujo indicando la forma de ejecutar la simulación.
 - (b) Escriba un programa que genere k ejecuciones.
 - (c) Verifique su programa comparando la salida con un cálculo a mano.
 - (d) Si $\alpha = 0.5, C = 5$ y $k = 100$, ejecute su programa y tome la salida para estimar el tiempo esperado hasta que falle el sistema.
14. Ciertos mensajes llegan a una instalación de comunicaciones de acuerdo con un proceso Poisson con razón de 2 por hora. La instalación consta de tres canales, cada mensaje llega a un canal libre, si los tres están libres, o se pierde si todos los canales están ocupados. El tiempo que los mensajes permanecen en un canal es una variable aleatoria que depende de las condiciones meteorológicas al momento de llegada. Específicamente, si el mensaje llega cuando las condiciones son "buenas",

entonces su tiempo de procesamiento es una variable aleatoria con función de distribución

$$F(x) = x, \quad 0 < x < 1$$

mientras que si las condiciones son "malas", su tiempo de procesamiento tiene la función de distribución

$$F(x) = x^3, \quad 0 < x < 1$$

Al principio, las condiciones son buenas, y alternan en periodos buenos y malos: los buenos tienen una duración fija de 2 horas y los malos de una (así, por ejemplo, en el instante 5, las condiciones cambian de buenas a malas).

Suponga que estamos interesados en la distribución del número de mensajes perdidos hasta el instante $T = 100$.

- (a) Defina los eventos y las variables que nos permitan utilizar el método de eventos discretos.
- (b) Escriba un diagrama de flujo de lo anterior.
- (c) Escriba un programa para lo anterior.
- (d) Verifique su programa comparando con una salida calculada a mano.
- (e) Ejecute su programa para estimar el número promedio de mensajes perdidos en las primeras 100 horas de operación.

15. Mediante un estudio de simulación, estime el valor esperado de una opción para comprar una acción en cualquier momento dentro de los próximos 20 días por un precio de 100, si el precio actual de la acción es 100. Suponga el modelo de la sección 6.8, con $\mu = -0.05, \sigma = 0.3$ y utilice la estrategia ahí mencionada.

Bibliografía

- Banks, J. y J. Carson, *Discrete-Event System Simulation*, Nueva Jersey, Prentice-Hall, 1984.
- Clymer, J., *Systems Analysis using Simulation and Markov Models*, Nueva Jersey, Prentice-Hall, 1990.
- Gottfried, B., *Elements of Stochastic Process Simulation*, Nueva Jersey, Prentice-Hall, 1984.
- Law, A. M., y W. D. Kelton, *Simulation Modelling and Analysis*, Nueva York, McGraw-Hill, 2a. ed., 1991.
- Mitrani, I., *Simulation Techniques for Discrete Event Systems*, Cambridge, Reino Unido, Cambridge University Press, 1982.
- Peterson, R. y E. Silver, *Decision Systems for Inventory Management and Production Planning*, Nueva York, Wiley, 1979.

Pritsker, A. y C. Pedgen, *Introduction to Simulation and SLAM*, Nueva York, Halsted Press, 1979.

Shannon, R. E., *Systems Simulation: The Art and Science*, Nueva Jersey, Prentice-Hall, 1975.

Solomon, S. L., *Simulation of Waiting Line Systems*, Nueva Jersey, Prentice-Hall, 1983.

CAPÍTULO.

7

Análisis estadístico
de datos simulados**Introducción**

Por lo general, los estudios de simulación se realizan para determinar el valor de cierta cantidad θ relacionada con un modelo estocástico particular. Una simulación del sistema en cuestión produce los datos de salida X , una variable aleatoria cuyo valor esperado es la cantidad de interés θ . Una simulación independiente (es decir, otra ejecución de la simulación) proporciona una nueva variable aleatoria independiente de la anterior, con media θ . Esto continúa hasta un total de k ejecuciones y k variables aleatorias independientes X_1, \dots, X_k , todas con la misma distribución y media θ . El promedio de estos k valores, $\bar{X} = \sum_{i=1}^k X_i / k$ sirve entonces como estimador, o aproximador, de θ .

En este capítulo consideramos el problema de decidir cuándo detener el estudio de simulación; es decir, el problema de hallar el valor adecuado de k . Para decidir cuándo detenerse, será útil considerar la calidad de nuestro estimador de θ . Además, mostraremos la forma de obtener un intervalo donde podamos afirmar que θ está ahí con cierto grado de confianza.

Las dos últimas secciones del capítulo muestran la forma de determinar la calidad de estimadores más complejos que la media muestral, utilizando una importante técnica estadística llamada "estimadores bootstrap". La aplicación del método bootstrap a los estudios de simulación aparece en la última sección del capítulo.

7.1 Media y varianza muestrales

Suponga que X_1, \dots, X_n son variables aleatorias independientes con la misma función de distribución. Sean θ y σ^2 su media y varianza, respectivamente; es decir, $\theta = E[X_i]$ y $\sigma^2 = \text{Var}(X_i)$. La cantidad

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

que es el promedio aritmético de los n datos es la *media muestral*. Cuando no se conoce la media θ poblacional, con frecuencia se toma la media muestral para estimarla.

Puesto que

$$\begin{aligned} E[\bar{X}] &= E\left[\sum_{i=1}^n \frac{X_i}{n}\right] \\ &= \sum_{i=1}^n \frac{E[X_i]}{n} \\ &= \frac{n\theta}{n} = \theta \end{aligned} \quad (7.1)$$

se tiene que \bar{X} es un estimador insesgado de θ ; decimos que un estimador de un parámetro es un estimador insesgado de ese parámetro si su valor esperado es igual al parámetro.

Para determinar la “bondad” de \bar{X} como estimador de la media poblacional θ , consideremos su error cuadrático medio; es decir, el valor esperado del cuadrado de la diferencia entre \bar{X} y θ . Ahora,

$$\begin{aligned} E[(\bar{X} - \theta)^2] &= \text{Var}(\bar{X}) \quad (\text{donde } E[\bar{X}] = \theta) \\ &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \quad (\text{es independiente}) \\ &= \frac{\sigma^2}{n} \quad (\text{pues } \text{Var}(X_i) = \sigma^2) \end{aligned} \quad (7.2)$$

Así, \bar{X} , la media muestral de los n valores X_1, \dots, X_n es una variable aleatoria con media θ y varianza σ^2/n . Como es poco probable que una variable aleatoria esté a una

distancia de muchas desviaciones estándar (la desviación estándar es la raíz cuadrada de su varianza) de su media, se tiene que \bar{X} es un buen estimador de θ cuando σ/\sqrt{n} es pequeño.

Observación La justificación de la afirmación anterior, que es poco probable que una variable aleatoria quede a demasiadas desviaciones estándar de su media, es una consecuencia de la desigualdad de Chebyshev y, algo más importante para los estudios de simulación, del teorema central de límite. De hecho, para cualquier $c > 0$, la desigualdad de Chebyshev (véase la sección 2.7 del capítulo 2) produce la cota conservadora

$$P\left\{|\bar{X} - \theta| > \frac{c\sigma}{\sqrt{n}}\right\} \leq \frac{1}{c^2}$$

Sin embargo, cuando n es grande, como es usual en las simulaciones, podemos aplicar el teorema central de límite para afirmar que $(\bar{X} - \theta)/(\sigma/\sqrt{n})$ se distribuye aproximadamente como una variable aleatoria normal unitaria, de modo que

$$\begin{aligned} P\{|\bar{X} - \theta| > c\sigma/\sqrt{n}\} &\approx P\{|Z| > c\}, \quad \text{donde } Z \text{ es una normal unitaria} \\ &= 2[1 - \phi(c)] \end{aligned} \quad (7.3)$$

donde ϕ es la función de distribución normal unitaria. Por ejemplo, como $\phi(1.96) \approx 0.975$, la ecuación (7.3) establece que la probabilidad de que la media muestral difiera de θ por más de $1.96\sigma/\sqrt{n}$ es aproximadamente 0.05, mientras que la desigualdad de Chebyshev, más débil, sólo implica que esta probabilidad es menor que $1/(1.96)^2 = 0.2603$. ■

La dificultad del uso directo de σ^2/n como indicación de lo bien o mal que la media muestral de n datos estima la media poblacional es que por lo general no se conoce la varianza poblacional σ^2 . Así, también necesitamos estimarla. Como

$$\sigma^2 = E[(X - \theta)^2]$$

es el promedio del cuadrado de la diferencia entre un dato y su media (desconocida), podría parecer, al tomar \bar{X} como estimador de la media, que un estimador natural de σ^2 sería $\sum_{i=1}^n (X_i - \bar{X})^2/n$, el promedio de los cuadrados de las distancias entre los datos y la media estimada. Sin embargo, para que el estimador sea insesgado (y por otras razones técnicas) preferimos dividir la suma de cuadrados entre $n - 1$ en vez de n .

Definición La cantidad S^2 , definida como

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

es la varianza muestral.

La identidad algebraica

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \quad (7.4)$$

cuya demostración se deja como ejercicio, permite mostrar que la varianza muestral es un estimador insesgado de σ^2 .

Proposición

$$E[S^2] = \sigma^2$$

Demostración Nos servimos de la identidad (7.4) para ver que

$$\begin{aligned} (n-1)E[S^2] &= E\left[\sum_{i=1}^n X_i^2\right] - nE[\bar{X}^2] \\ &= nE[X_1^2] - nE[\bar{X}^2] \end{aligned} \quad (7.5)$$

donde la última igualdad se debe a que todas las X_i tienen la misma distribución. Si recordamos que para cualquier variable aleatoria Y , $\text{Var}(Y) = E[Y^2] - (E[Y])^2$ o, en forma equivalente,

$$E[Y^2] = \text{Var}(Y) + (E[Y])^2$$

obtenemos que

$$\begin{aligned} E[X_1^2] &= \text{Var}(X_1) + (E[X_1])^2 \\ &= \sigma^2 + \theta^2 \end{aligned}$$

y

$$\begin{aligned} E[\bar{X}^2] &= \text{Var}(\bar{X}) + (E[\bar{X}])^2 \\ &= \frac{\sigma^2}{n} + \theta^2 \quad [\text{de (7.2) y (7.1)}] \end{aligned}$$

Así, de la ecuación (7.5), obtenemos que

$$(n-1)E[S^2] = n(\sigma^2 + \theta^2) - n\left(\frac{\sigma^2}{n} + \theta^2\right) = (n-1)\sigma^2$$

lo cual demuestra el resultado. ■

Empleamos la varianza muestral S^2 como nuestro estimador de la varianza poblacional σ^2 y a $S = \sqrt{S^2}$, la llamada desviación estándar muestral, como estimador de σ .

Ahora, supongamos que, como en una simulación, tenemos la opción de generar de manera continua más datos X_i . Si nuestro objetivo es estimar el valor de $\theta = E[X_i]$, ¿en qué momento debemos detenernos al ir generando nuevos datos? La respuesta es que primero debemos elegir un valor aceptable d para la desviación estándar de nuestro estimador, ya que si d es la desviación estándar del estimador \bar{X} , entonces podemos, por ejemplo, estar 95 por ciento seguros de que \bar{X} no difiere de θ en más de $1.96d$. Por lo tanto, debemos continuar generando nuevos datos hasta generar n datos para los cuales nuestra estimación de σ/\sqrt{n} (a saber, S/\sqrt{n}) sea menor que un valor aceptable d . Como la desviación estándar muestral S podría no ser una buena estimación particular de σ (incluso podría no ser válida la aproximación normal) cuando el tamaño de la muestra es pequeño, recomendamos el siguiente procedimiento para determinar el momento de detenernos.

Método para determinar cuándo detenerse al generar nuevos datos

1. Elegir un valor aceptable d para la desviación estándar del estimador.
2. Generar al menos 30 datos.
3. Continuar generando más datos, hasta detenerse cuando se hayan generado k valores y $S/\sqrt{k} < d$, donde S es la desviación estándar muestral basada en estos k valores.
4. La estimación de θ está dada por $\bar{X} = \sum_{i=1}^k X_i/k$.

Ejemplo 7a Considere un sistema de servicio, donde no se permite la entrada de nuevos clientes después de las cinco de la tarde. Suponga que cada día sigue la misma ley de probabilidad y que estamos interesados en estimar la hora esperada en que el último cliente sale del sistema. Además, suponga que queremos estar 95 por ciento seguros de que nuestra respuesta estimada no diferirá del valor real en más de 15 segundos.

Para satisfacer el requisito anterior, es necesario que generemos de manera continua datos relacionados con la hora en que sale el último cliente (cada vez mediante una ejecución de simulación) hasta haber generado un total de k valores, donde k es al

menos 30 y es tal que $1.96S/\sqrt{k} < 15$, donde S es la desviación estándar muestral (medida en segundos) de estos k datos. Nuestra estimación de la hora esperada en que saldrá el último cliente será el promedio de los k datos. ■

Para utilizar la técnica anterior y determinar cuándo dejar de generar nuevos valores, sería útil tener un método para calcular de manera recursiva las sucesivas medias y varianzas muestrales, en vez de calcular todo desde el principio cada vez que se genera un nuevo dato. Ahora mostraremos cómo hacerlo. Consideremos la sucesión de datos X_1, X_2, \dots y sean

$$\bar{X}_j = \sum_{i=1}^j \frac{X_i}{j}$$

y

$$S_j^2 = \sum_{i=1}^j \frac{(X_i - \bar{X}_j)^2}{j-1}, \quad j \geq 2$$

la media y la varianza muestrales, respectivamente, de los primeros j datos. Las siguientes fórmulas recursivas permiten calcular de manera sucesiva el valor actual de la media y de la varianza muestrales.

Con $S_1^2 = 0$, $\bar{X}_0 = 0$.

$$\bar{X}_{j+1} = \bar{X}_j + \frac{X_{j+1} - \bar{X}_j}{j+1} \quad (7.6)$$

$$S_{j+1}^2 = \left(1 - \frac{1}{j}\right) S_j^2 + (j+1)(\bar{X}_{j+1} - \bar{X}_j)^2 \quad (7.7)$$

Ejemplo 7b Si los primeros tres datos son $X_1 = 5$, $X_2 = 14$, $X_3 = 9$, entonces las ecuaciones (7.6) y (7.7) implican que

$$\bar{X}_1 = 5$$

$$\bar{X}_2 = 5 + \frac{9}{2} = \frac{19}{2}$$

$$S_2^2 = 2\left(\frac{19}{2} - 5\right)^2 = \frac{81}{2}$$

$$\bar{X}_3 = \frac{19}{2} + \frac{1}{3}\left(9 - \frac{19}{2}\right) = \frac{28}{3}$$

$$S_3^2 = \frac{81}{4} + 3\left(\frac{28}{3} - \frac{19}{2}\right)^2 = \frac{61}{3}$$

■

El análisis se modifica un poco cuando los datos son variables aleatorias Bernoulli (o variables aleatorias 0-1), como cuando estimamos una probabilidad. Es decir, suponga que podemos generar variables aleatorias X tales que

$$X_i = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p \end{cases}$$

y suponga que estamos interesados en estimar $E[X_i] = p$. Como en esta situación se tiene que

$$\text{Var}(X_i) = p(1-p)$$

no se necesita utilizar la varianza muestral para estimar $\text{Var}(X_i)$. De hecho, si hemos generado n datos X_1, \dots, X_n , entonces, como la estimación de p será

$$\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}$$

una estimación natural de $\text{Var}(X_i)$ es $\bar{X}_n(1 - \bar{X}_n)$. Por lo tanto, en este caso, tenemos el siguiente método para decidir cuándo detenernos.

1. Elegir un valor aceptable d para la desviación estándar del estimador.
2. Generar al menos 30 valores de datos.
3. Continuar generando más datos, hasta detenerse al generar k valores y $[\bar{X}_k(1 - \bar{X}_k)/k]^{1/2} < d$.
4. La estimación de p es \bar{X}_k , el promedio de los k valores de datos.

Ejemplo 7c En el ejemplo 7a, suponga que estamos interesados en estimar la probabilidad de que aún haya un cliente en la tienda a las cinco y media. Para esto, simularíamos varios días sucesivos, y sea

$$X_i = \begin{cases} 1 & \text{si hay un cliente presente a las cinco y media del día } i \\ 0 & \text{en caso contrario} \end{cases}$$

Simularíamos al menos 30 días y continuaríamos simulando hasta el k -ésimo día, donde k es tal que $[p_k(1 - p_k)/k]^{1/2} < d$, donde $p_k = \bar{X}_k$ es la proporción de estos k días en los que hay un cliente presente a las cinco y media y donde d es un valor aceptable para la desviación estándar del estimador p_k . ■

7.2 Estimación del intervalo de una media poblacional

Suponga de nuevo que X_1, X_2, \dots, X_n son variables aleatorias independientes con una distribución común, con media θ y varianza σ^2 . Aunque la media muestral

$\bar{X} = \sum_{i=1}^n X_i/n$ es un estimador eficaz de θ , no esperamos que \bar{X} sea igual a θ , sino que sea “cercano”. Como resultado, a veces es más útil especificar un intervalo para el cual tenemos cierto grado de confianza de que θ esté en él.

Para obtener tal intervalo, necesitamos la distribución (aproximada) del estimador \bar{X} . Para esto, primero recordemos, de las ecuaciones (7.1) y (7.2), que

$$E[\bar{X}] = \theta, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

y así, por el teorema central de límite, se tiene que para n grande

$$\sqrt{n} \frac{(\bar{X} - \theta)}{\sigma} \sim N(0, 1)$$

donde $\sim N(0, 1)$ significa “se distribuye aproximadamente como una normal unitaria”. Además, si reemplazamos la desviación estándar desconocida σ por su estimador S , la desviación estándar muestral, entonces sigue ocurriendo (por un resultado conocido como teorema de Slutsky) que la cantidad resultante es aproximadamente una normal unitaria. Es decir, cuando n es grande,

$$\sqrt{n}(\bar{X} - \theta)/S \sim N(0, 1) \quad (7.8)$$

Ahora, para cualquier α , $0 < \alpha < 1$, sea Z_α tal que

$$P\{Z > z_\alpha\} = \alpha$$

donde Z es una variable aleatoria normal unitaria (por ejemplo, $z_{0.05} = 1.96$). La simetría de la función de densidad normal unitaria con respecto del origen implica que $z_{1-\alpha}$, el punto donde el área debajo de la densidad y a la derecha del punto es igual a $1 - \alpha$, es tal que (véase la figura 7.1)

$$z_{1-\alpha} = -z_\alpha$$

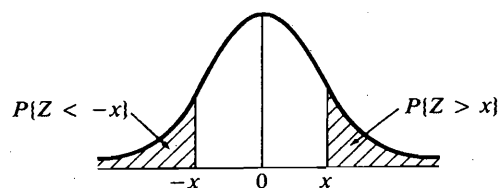


Figura 7.1 Densidad normal unitaria.

Por lo tanto (véase la figura 7.1)

$$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$$

Así, (7.8) implica que

$$P\left\{-z_{\alpha/2} < \sqrt{n} \frac{(\bar{X} - \theta)}{S} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

o en forma equivalente, multiplicando por -1 ,

$$P\left\{-z_{\alpha/2} < \sqrt{n} \frac{(\theta - \bar{X})}{S} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

que es equivalente a

$$P\left\{\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right\} \approx 1 - \alpha \quad (7.9)$$

En otras palabras, con probabilidad $1 - \alpha$, la media poblacional θ estará dentro de la región $\bar{X} \pm z_{\alpha/2} S/\sqrt{n}$.

Definición Si los valores observados de la media y la desviación estándar muestrales son $\bar{X} = \bar{x}$ y $S = s$, respectivamente, el intervalo $\bar{x} \pm z_{\alpha/2} s/\sqrt{n}$ es una estimación de θ mediante un intervalo de confianza del 100 $(1 - \alpha)$ por ciento (aproximadamente).

Observaciones

1. Para aclarar el significado de un “intervalo de confianza del 100 $(1 - \alpha)$ por ciento”, consideremos, por ejemplo, el caso en que $\alpha = 0.05$, de modo que $z_{\alpha/2} = 1.96$. Ahora, antes de observar los datos, será cierto, con probabilidad (aproximadamente) igual a 0.95, que la media muestral \bar{X} y la desviación estándar muestral S serán tales que θ estará entre $\bar{X} \pm 1.96 S/\sqrt{n}$. Después de observar que \bar{X} y S sean iguales a \bar{x} y s , respectivamente, ya no hay una probabilidad de que θ esté en el intervalo $\bar{x} \pm 1.96 s/\sqrt{n}$, pues ahora está o no está ahí. Sin embargo, estamos “95 por ciento seguros” de que en esta situación está dentro del intervalo (pues sabemos que, a largo plazo, tales intervalos contendrán la media 95 por ciento de las veces).
2. (Una observación técnica.) El análisis anterior se basa en la ecuación (7.8), la cual establece que $\sqrt{n}(\bar{X} - \theta)/S$ es aproximadamente una variable aleatoria normal unitaria cuando n es grande. Ahora, si los datos originales X_i tuvieran a su vez una distribución normal, entonces se sabe que esta cantidad tiene (exactamente) una distribución t con $n - 1$ grados de libertad. Por esta razón, muchos

autores han propuesto esta distribución aproximada en el caso general en que la distribución original no necesariamente sea normal. Sin embargo, como no es claro que la distribución t con $n - 1$ grados de libertad sea una mejor aproximación que la normal en el caso general y como estas distribuciones son aproximadamente iguales para n grande, hemos empleado la aproximación normal en vez de presentar la variable aleatoria t . ■

Consideremos ahora el caso, como en un estudio de simulación, en que se pueden generar más datos y la cuestión consiste en determinar cuándo dejar de generarlos. Una solución consiste en elegir valores iniciales α y l y continuar generando datos hasta que la estimación de θ mediante un intervalo de confianza de $100(1 - \alpha)$ por ciento sea menor que l . Como la longitud de este intervalo será $2z_{\alpha/2}S/\sqrt{n}$, podemos lograrlo mediante la técnica siguiente.

1. Generar al menos 30 valores de datos.
2. Continuar generando más datos y detenerse cuando el número de valores generados (llamémosle k) es tal que $2z_{\alpha/2}S/\sqrt{k} < l$, donde S es la desviación estándar muestral basada en esos k valores [el valor de S debe actualizarse constantemente, mediante la recursión dada por (7.6) y (7.7), al irse generando nuevos datos].
3. Si \bar{x} y s son los valores observados de \bar{X} y S , entonces la estimación de θ mediante un intervalo de confianza de $100(1 - \alpha)$ por ciento, cuya longitud es menor que l , es $\bar{x} \pm 2z_{\alpha/2}s/\sqrt{k}$.

Una observación técnica Un lector con mayores conocimientos de estadística podría dudar de nuestro uso de un intervalo de confianza aproximado, cuya teoría se basa en la hipótesis de que el tamaño de la muestra era fijo, cuando en la situación anterior tal tamaño es sin duda una variable aleatoria que depende de los datos generados. Sin embargo, esto se justifica cuando el tamaño muestral es grande, de modo que desde el punto de vista de la simulación podemos ignorar esta sutileza. ■

Como se observó en la sección anterior, el análisis se modifica cuando X_1, \dots, X_n son variables aleatorias Bernoulli tales que

$$X_i = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p \end{cases}$$

Como en este caso podemos estimar $\text{Var}(X_i)$ mediante $\bar{X}(1 - \bar{X})$, tenemos que la afirmación equivalente a la ecuación (7.8) es que cuando n es grande,

$$\sqrt{n} \frac{(\bar{X} - p)}{\sqrt{\bar{X}(1 - \bar{X})}} \sim N(0, 1) \quad (7.10)$$

Por lo tanto, para cualquier α ,

$$P\left\{-z_{\alpha/2} < \sqrt{n} \frac{(\bar{X} - p)}{\sqrt{\bar{X}(1 - \bar{X})}} < z_{\alpha/2}\right\} = 1 - \alpha$$

o, en forma equivalente,

$$P(\bar{X} - z_{\alpha/2}\sqrt{\bar{X}(1 - \bar{X})/n} < p < \bar{X} + z_{\alpha/2}\sqrt{\bar{X}(1 - \bar{X})/n}) = 1 - \alpha$$

Por lo tanto, si el valor observado de \bar{X} es p_n , decimos que la “estimación de p mediante un intervalo de confianza de $100(1 - \alpha)$ por ciento” es

$$p_n \pm z_{\alpha/2}\sqrt{p_n(1 - p_n)/n}$$

7.3 La técnica bootstrap para estimar errores cuadráticos medios

Suponga ahora que X_1, \dots, X_n son variables aleatorias independientes con una función de distribución común F y que estamos interesados en utilizarlas para estimar cierto parámetro $\theta(F)$ de la distribución F . Por ejemplo, $\theta(F)$ podría ser (como en las secciones anteriores del capítulo) la media, la mediana o la varianza de F o cualquier otro parámetro de F . Suponga además que se ha propuesto un estimador de $\theta(F)$, denotado $g(X_1, \dots, X_n)$, y para juzgar su valor como estimador de $\theta(F)$ queremos estimar su error cuadrático medio. Es decir, queremos estimar el valor de

$$\text{ECM}(F) \equiv E_F[(g(X_1, \dots, X_n) - \theta(F))^2]$$

[nuestra elección de la notación $\text{ECM}(F)$ suprime la dependencia del estimador g ; usamos la notación E_F para indicar que la esperanza debe tomar en cuenta la hipótesis de que todas las variables aleatorias tienen la distribución F]. Ahora, aunque hay un estimador inmediato de este ECM (a saber, S^2/n) cuando $\theta(F) = E[X_i]$ y $g(X_1, \dots, X_n) = \bar{X}$, no es evidente la forma de estimarlo en otros casos. Ahora presentamos una técnica útil, conocida como la técnica bootstrap, para estimar este error cuadrático medio.

Para comenzar, observemos que si conociéramos la función de distribución F , entonces podríamos calcular teóricamente la esperanza del cuadrado de la diferencia entre θ y su estimador; es decir, podríamos calcular el error cuadrático medio. Sin embargo, después de observar los valores de los n datos, tenemos una buena idea de la distribución subyacente. De hecho, suponga que los valores observados de los datos

son $X_i = x_i$, $i = 1, \dots, n$. Ahora podemos estimar la función de distribución subyacente F mediante la llamada función de distribución empírica F_e , donde $F_e(x)$, la estimación de $F(x)$, la probabilidad de que un dato sea menor o igual a x , es precisamente la proporción de los n datos que son menores o iguales a x . Es decir,

$$F_e(x) = \frac{\text{número de } i : X_i \leq x}{n}$$

Otra forma de pensar acerca de F_e es que es la función de distribución de una variable aleatoria X_e , la cual tiene la misma probabilidad de tomar cualquiera de los n valores x_i , $i = 1, \dots, n$ (si los valores x_i no son todos distintos, entonces interpretamos lo anterior diciendo que X_e será igual al valor x_i con una probabilidad igual al número de j tales que $x_j = x_i$ dividido entre n ; es decir, si $n = 3$ y $x_1 = x_2 = 1$, $x_3 = 2$, entonces X_e es una variable aleatoria que toma el valor 1 con probabilidad $\frac{2}{3}$ y 2 con probabilidad $\frac{1}{3}$).

Ahora, si F_e está "cerca" de F , como debería ser si n es grande [de hecho, la ley fuerte de los grandes números implica que con probabilidad 1, $F_e(x)$ converge a $F(x)$ cuando $n \rightarrow \infty$; otro resultado, conocido como el teorema de Glivenko-Cantelli, establece que, con probabilidad 1, esta convergencia será uniforme en x], entonces $\theta(F_e)$ estará probablemente cerca de $\theta(F)$ (suponiendo que θ es, en cierto sentido, una función continua de la distribución) y el $\text{ECM}(F)$ sería aproximadamente igual a

$$\text{ECM}(F_e) = E_{F_e}[(g(X_1, \dots, X_n) - \theta(F_e))^2]$$

En la expresión anterior, los X_i se consideran como variables aleatorias independientes con función de distribución F_e . La cantidad $\text{ECM}(F_e)$ es la *aproximación bootstrap del error cuadrático medio* $\text{ECM}(F)$.

Para tener una idea de la eficacia de la aproximación bootstrap del error cuadrático medio, consideremos el caso en que su uso no es necesario, a saber, la estimación de la media de una distribución mediante la media muestral \bar{X} . (En este caso, su uso no es necesario, pues ya hay una forma eficaz de estimar el error cuadrático medio $E[(\bar{X} - \theta)^2] = \sigma^2/n$, que consiste en utilizar el valor observado de S^2/n .)

Ejemplo 7d Suponga que estamos interesados en estimar $\theta(F) = E[X]$ mediante la media muestral $\bar{X} = \sum_{i=1}^n X_i/n$. Si los datos observados son x_i , $i = 1, \dots, n$, entonces la distribución empírica F_e coloca el peso $1/n$ en cada uno de los puntos x_1, \dots, x_n (combinando los pesos si las x_i no son todas distintas). Por lo tanto, la media de F_e es

$\theta(F_e) = \bar{x} = \sum_{i=1}^n X_i/n$, y entonces la estimación bootstrap del error cuadrático medio [denotada $\text{ECM}(F_e)$] está dada por

$$\text{ECM}(F_e) = E_{F_e} \left[\left(\sum_{i=1}^n \frac{X_i}{n} - \bar{x} \right)^2 \right]$$

donde X_1, \dots, X_n son variables aleatorias independientes, cada una distribuida de acuerdo con F_e . Como

$$E_{F_e} \left[\sum_{i=1}^n \frac{X_i}{n} \right] = E_{F_e}[X] = \bar{x}$$

se tiene que

$$\begin{aligned} \text{ECM}(F_e) &= \text{Var}_{F_e} \left(\sum_{i=1}^n \frac{X_i}{n} \right) \\ &= \frac{\text{Var}_{F_e}(X)}{n} \end{aligned}$$

Ahora,

$$\begin{aligned} \text{Var}_{F_e}(X) &= E_{F_e}[(X - E_{F_e}[X])^2] \\ &= E_{F_e}[(X - \bar{x})^2] \\ &= \frac{1}{n} \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \end{aligned}$$

y entonces

$$\text{ECM}(F_e) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n^2}$$

lo cual se equipara bastante bien con S^2/n , la estimación usual del error cuadrático medio. De hecho, como el valor observado de S^2/n es $\sum_{i=1}^n (x_i - \bar{x})^2/[n(n-1)]$, la aproximación bootstrap es casi idéntica. ■

Si los datos son $X_i = x_i$, $i = 1, \dots, n$, entonces, como la función de distribución empírica F_e coloca pesos $1/n$ en cada uno de los puntos x_i , por lo general es fácil calcular el valor de $\theta(F_e)$: por ejemplo, si el parámetro de interés $\theta(F)$ era la varianza

de la distribución F , entonces $\theta(F_e) = \text{Var}_{F_e}(X) = \sum_{i=1}^n (x_i - \bar{x})^2/n$. Para determinar la aproximación bootstrap del error cuadrático medio, tenemos que calcular

$$\text{ECM}(F_e) = E_{F_e}[g(X_1, \dots, X_n) - \theta(F_e)]^2$$

Sin embargo, como esta esperanza debe calcularse según la hipótesis de que X_1, \dots, X_n son variables aleatorias independientes distribuidas de acuerdo con F_e , implica que el vector (X_1, \dots, X_n) tiene la misma probabilidad de tomar cualquiera de los n^n valores posibles $(x_{i1}, x_{i2}, \dots, x_{in})$, $ij \in \{1, 2, \dots, n\}$, $j = 1, \dots, n$. Por lo tanto,

$$\text{ECM}(F_e) = \sum_{i_n} \dots \sum_{i_1} \frac{[g(x_{i1}, \dots, x_{in}) - \theta(F_e)]^2}{n^n}$$

donde cada i_j va de 1 a n , de modo que el cálculo del $\text{ECM}(F_e)$ requiere, en general, la suma de n^n términos, una tarea imposible si n es grande.

Sin embargo, como sabemos, hay una forma eficaz de determinar el promedio de un gran número de términos, a saber, mediante simulación. De hecho, podríamos generar un conjunto de n variables aleatorias independientes X_1^1, \dots, X_n^1 , cada una con función de distribución F_e , y luego hacer

$$Y_1 = [g(X_1^1, \dots, X_n^1) - \theta(F_e)]^2$$

Luego generamos un segundo conjunto X_1^2, \dots, X_n^2 y calculamos

$$Y_2 = [g(X_1^2, \dots, X_n^2) - \theta(F_e)]^2$$

y así sucesivamente, hasta obtener las variables Y_1, Y_2, \dots, Y_r . Como estas Y_i son variables aleatorias independientes con media $\text{ECM}(F_e)$, esto implica que podemos utilizar su promedio $\sum_{i=1}^r Y_i/r$ como una estimación del $\text{ECM}(F_e)$.

Observaciones

1. Es fácil generar una variable aleatoria X con distribución F_e . Como esta variable aleatoria debe tomar con la misma probabilidad a x_1, \dots, x_n , basta generar un número aleatorio U y hacer $X = x_I$, donde $I = \text{Ent}(nU) + 1$ (es fácil verificar que esto funciona incluso aunque los x_i no sean todos distintos).
2. La simulación anterior permite aproximar $\text{ECM}(F_e)$, que es a su vez una aproximación del deseado $\text{ECM}(F)$. Como tal, se ha indicado que por lo general basta cerca de 100 ejecuciones de simulación; es decir, basta elegir $r = 100$. ■

El siguiente ejemplo ilustra el uso del método bootstrap para analizar la salida de la simulación de una cola.

Ejemplo 7e En el ejemplo 7a, suponga que estamos interesados en estimar el tiempo promedio a largo plazo que un cliente está en el sistema. Es decir, si W_i es el tiempo que el cliente i -ésimo ocupa en el sistema, $i \geq 1$, estamos interesados en

$$\theta \equiv \lim_{n \rightarrow \infty} \frac{W_1 + W_2 + \dots + W_n}{n}$$

Para mostrar que este límite existe (observe que las variables aleatorias W_i no son independientes ni idénticamente distribuidas; ¿por qué?), sea N_i el número de clientes que llegan el día i , y sean

$$D_1 = W_1 + \dots + W_{N_1}$$

$$D_2 = W_{N_1+1} + \dots + W_{N_1+N_2}$$

y en general, para $i > 2$,

$$D_i = W_{N_1+\dots+N_{i-1}+1} + \dots + W_{N_1+\dots+N_i}$$

En palabras, D_i es la suma de las horas en el sistema para todas las llegadas en el día i . Ahora podemos expresar θ como

$$\theta = \lim_{m \rightarrow \infty} \frac{D_1 + D_2 + \dots + D_m}{N_1 + N_2 + \dots + N_m}$$

donde lo anterior es consecuencia del hecho de que el cociente no es más que el tiempo promedio en el sistema de todos los clientes que llegan en los primeros m días. Al dividir el numerador y el denominador entre m , obtenemos

$$\theta = \lim_{m \rightarrow \infty} \frac{(D_1 + \dots + D_m)/m}{(N_1 + \dots + N_m)/m}$$

Ahora, como cada día sigue la misma ley de probabilidad, se tiene que las variables aleatorias D_1, \dots, D_m son todas independientes e idénticamente distribuidas, al igual que lo son las variables aleatorias N_1, \dots, N_m . Por lo tanto, por la ley fuerte de los grandes números, se tiene que el promedio de los primeros m de los D_i convergerá,

con probabilidad 1, a su esperanza común, y se cumplirá una afirmación similar para los N_i . Por lo tanto, vemos que

$$\theta = \frac{E[D]}{E[N]}$$

donde $E[N]$ es la cantidad esperada de clientes en un día, y $E[D]$ es la suma esperada de las horas que los clientes pasan en el sistema.

Entonces, para estimar θ podemos simular el sistema durante k días, reuniendo en la i -ésima ejecución los datos N_i, D_i , donde N_i es la cantidad de clientes que llegan el día i y D_i es la suma de las horas que pasan en el sistema, $i = 1, \dots, k$. Como la cantidad $E[D]$ se puede estimar ahora mediante

$$\bar{D} = \frac{D_1 + D_2 + \dots + D_k}{k}$$

y $E[N]$ mediante

$$\bar{N} = \frac{N_1 + N_2 + \dots + N_k}{k}$$

se tiene que $\theta = E[D]/E[N]$ se puede estimar como

$$\text{Estimación de } \theta = \frac{\bar{D}}{\bar{N}} = \frac{D_1 + \dots + D_k}{N_1 + \dots + N_k}$$

lo cual, debemos observar, es simplemente el tiempo promedio en el sistema de todas las llegadas ocurridas durante los primeros k días.

Para estimar

$$\text{ECM} = E \left[\left(\frac{\sum_{i=1}^k D_i}{\sum_{i=1}^k N_i} - \theta \right)^2 \right]$$

utilizamos el método bootstrap. Supongamos que los valores observados de D_i y N_i son d_i y n_i , $i = 1, \dots, k$, es decir, que la simulación produce n_i llegadas el día i y el sistema demora un tiempo total d_i . Así, la función de distribución conjunta empírica del vector aleatorio (D, N) coloca el mismo peso en las k parejas (d_i, n_i) , $i = 1, \dots, k$. Es decir, bajo la función de distribución empírica tenemos

$$P_{F_e}\{D = d_i, N = n_i\} = \frac{1}{k}, \quad i = 1, \dots, k$$

Por lo tanto,

$$E_{F_e}[D] = \bar{d} = \frac{\sum_{i=1}^k d_i}{k}, \quad E_{F_e}[N] = \bar{n} = \frac{\sum_{i=1}^k n_i}{k}$$

y así,

$$\theta(F_e) = \frac{\bar{d}}{\bar{n}}$$

Por lo tanto,

$$\text{ECM}(F_e) = E_{F_e} \left[\left(\frac{\sum_{i=1}^k D_i}{\sum_{i=1}^k N_i} - \frac{\bar{d}}{\bar{n}} \right)^2 \right]$$

donde lo anterior se calcula bajo la hipótesis de que las k parejas de vectores aleatorios (D_i, N_i) se distribuyen de manera independiente de acuerdo con F_e .

Como el cálculo exacto del $\text{ECM}(F_e)$ requeriría el cálculo de la suma de k^2 términos, ahora realizaremos un experimento de simulación para aproximarlos. Generamos k parejas independientes de vectores aleatorios (D_i^1, N_i^1) , $i = 1, \dots, k$, de acuerdo con la función de distribución empírica F_e , y luego calculamos

$$Y_1 = \left(\frac{\sum_{i=1}^k D_i^1}{\sum_{i=1}^k N_i^1} - \frac{\bar{d}}{\bar{n}} \right)^2$$

A continuación generamos un segundo conjunto (D_i^2, N_i^2) y calculamos la Y_2 correspondiente. Esto continúa hasta generar los r valores Y_1, \dots, Y_r (donde $r = 100$ bastaría). El promedio de estos r valores, $\sum_{i=1}^r Y_i / r$ se utiliza entonces para estimar $\text{ECM}(F_e)$, que es a su vez nuestra estimación del ECM, el error cuadrático medio de nuestra estimación de la cantidad promedio de tiempo que un cliente pasa en el sistema. ■

Observación: el método regenerativo En el análisis anterior se supuso que cada día seguía de manera independiente las mismas leyes de probabilidad. En ciertas aplicaciones, las mismas leyes de probabilidad describen al sistema no durante días de longitud fija, sino en ciclos con longitud aleatoria. Por ejemplo, consideremos un sistema de línea de espera en el que los clientes llegan de acuerdo con un proceso

Poisson y supongamos que el primer cliente llega en el instante 0. Si el tiempo aleatorio T representa el siguiente instante en que una llegada encuentra al sistema vacío, entonces decimos que el tiempo de 0 a T constituye el primer ciclo. El segundo ciclo sería entonces el tiempo desde T hasta el primer instante posterior a T en que una llegada encuentra al sistema vacío, y así sucesivamente. En la mayor parte de los modelos, es fácil ver que los movimientos del proceso en cada ciclo son independientes e idénticamente distribuidos. Por lo tanto, si consideramos un ciclo como un "día", entonces todo el análisis anterior sigue siendo válido. Por ejemplo, θ , el tiempo promedio que un cliente pasa en el sistema, está dada por $\theta = E[D]/E[N]$, donde D es la suma de los tiempos en el sistema de todas las llegadas en un ciclo y N es el número de tales llegadas. Si ahora generamos k ciclos, nuestra estimación de θ sigue siendo $\sum_{i=1}^k D_i / \sum_{i=1}^k N_i$. Además, el error cuadrático medio de esta estimación se puede aproximar mediante el método bootstrap, exactamente como antes.

La técnica de análisis de un sistema simulando "ciclos", es decir, intervalos aleatorios durante los cuales el proceso sigue la misma ley de probabilidad, se conoce como el método regenerativo. ■

Ejercicios

1. Para cualquier conjunto de números x_1, \dots, x_n , demuestre en forma algebraica que

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

donde $\bar{x} = \sum_{i=1}^n x_i / n$.

2. Dé una demostración probabilística del resultado del ejercicio 1, haciendo que X denote una variable aleatoria que, con la misma probabilidad, puede tomar los valores x_1, \dots, x_n , y luego aplique la identidad $\text{Var}(X) = E[X^2] - (E[X])^2$.
3. Escriba un programa que utilice las fórmulas recursivas dadas en las ecuaciones (7.6) y (7.7) para calcular la media y la varianza muestrales de un conjunto de datos.
4. Continúe generando variables aleatorias normales unitarias hasta generar n de ellas, donde $n \geq 30$ es tal que $S/\sqrt{n} < 0.1$, y S es la desviación estándar muestral de los n datos.
- ¿Cuántas normales cree que se generen?
 - ¿Cuántas normales generó?
 - ¿Cuál es la media muestral de todas las normales generadas?
 - ¿Cuál es la varianza muestral?
 - Comente los resultados de (c) y (d) ¿Son sorprendentes?

5. Repita el ejercicio 4, pero ahora continúe generando normales unitarias hasta que $S/\sqrt{n} < 0.01$.
6. Estime $\int_0^1 \exp(x^2) dx$ generando números aleatorios. Genere al menos 100 valores y deténgase cuando la desviación estándar de su estimador sea menor que 0.01.
7. Para estimar $E[X]$, se han simulado X_1, \dots, X_{16} con los siguientes valores resultantes: 10, 11, 10.5, 11.5, 14, 8, 13, 6, 15, 10, 11.5, 10.5, 12, 8, 16, 5. Con base en estos datos, si queremos que la desviación estándar del estimador de $E[X]$ sea menor que 0.1, ¿aproximadamente cuántas ejecuciones de simulación se necesitan?

Los ejercicios 8 y 9 se relacionan con la estimación de e .

8. Se puede mostrar que si sumamos números aleatorios hasta que su suma sea mayor que 1, entonces la cantidad esperada de sumandos es igual a e . Es decir, si

$$N = \min \left\{ n: \sum_{i=1}^n U_i > 1 \right\}$$

entonces $E[N] = e$.

- Sírvase de lo anterior para estimar e mediante 1000 ejecuciones de simulación.
 - Estime la varianza del estimador en (a) y dé una estimación de e mediante un intervalo de confianza de 95 por ciento.
9. Considere una sucesión de números aleatorios y sea M el primero que es menor que su predecesor. Es decir,
- $$M = \min \{ n: U_1 \leq U_2 \leq \dots \leq U_{n-1} > U_n \}$$
- Justifique que $P\{M > n\} = \frac{1}{n!}$, $n \geq 0$.
 - Utilice la identidad $E[M] = \sum_{n=0}^{\infty} P\{M > n\}$ para mostrar que $E[M] = e$.
 - Utilice la parte (b) para estimar e mediante 1000 ejecuciones de simulación.
 - Estime la varianza del estimador en (c) y dé una estimación de e mediante un intervalo de confianza de 95 por ciento.
10. Utilice el método del ejemplo 3a del capítulo 3 para obtener un intervalo de tamaño menor que 0.1, el cual contenga a π con 95 por ciento de confianza. ¿Cuántas ejecuciones fueron necesarias?
11. Repita el ejercicio 10, si queremos que el intervalo no sea mayor que 0.01.
12. Para estimar θ , generamos 20 valores independientes con media θ . Si los valores sucesivos obtenidos fueron

102,	112,	131,	107,	114,	95,	133,	145,	139,	117
93,	111,	124,	122,	136,	141,	119,	122,	151,	143

¿cuántas variables aleatorias adicionales cree que tengamos que generar si queremos tener 99 por ciento de seguridad de que nuestra estimación final de θ es correcta salvo un error de ± 0.5 ?

13. Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con media μ desconocida. Para a y b constantes dadas, $a < b$, nos interesa estimar $p = P\{\alpha < \Sigma_{i=1}^n x_i/n - \mu < b\}$.

- (a) Explique cómo podemos utilizar el método bootstrap para estimar a p .
 (b) Estime p si $n = 10$ y los valores de las X_i son 56, 101, 78, 67, 93, 87, 64, 72, 80 y 69. Sean $a = -5$, $b = 5$.

En los tres ejercicios siguientes, X_1, \dots, X_n es una muestra de una distribución cuya varianza (desconocida) es σ^2 . Planeamos estimar σ^2 mediante la varianza muestral $S^2 = \Sigma_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$, y queremos utilizar la técnica bootstrap para estimar $\text{Var}(S^2)$.

14. Si $n = 2$, $X_1 = 1$ y $X_2 = 3$, ¿cuál es la estimación bootstrap de $\text{Var}(S^2)$?

15. Si $n = 15$ y los datos son

5, 4, 9, 6, 21, 17, 11, 20, 7, 10, 21, 15, 13, 16, 8

aproxime (mediante una simulación) la estimación bootstrap de $\text{Var}(S^2)$.

16. Considere un sistema con un único servidor en el cual los clientes potenciales llegan de acuerdo con un proceso Poisson con razón 4.0. Un cliente potencial entrará al sistema sólo si hay tres o menos clientes en el sistema al momento de su llegada. El tiempo de servicio de cada uno es exponencial con razón 4.2. Después del instante $T = 8$ no entran más clientes al sistema (los tiempos están dados en horas). Realice un estudio de simulación para estimar el tiempo promedio que un cliente pasa en el sistema. Aplique el método bootstrap para estimar el error cuadrático medio de su estimador.

Bibliografía

- Bratley, P., B. L. Fox y L. E. Schrage, *A Guide to Simulation*, Nueva York, Springer Verlag, 2a. ed., 1988.
 Crane, M. A. y A. J. Lemoine, *An Introduction to the Regenerative Method for Simulation Analysis*, Nueva York, Springer Verlag, 1977.
 Efron, B. y R. Tibshirani, *Introduction to the Bootstrap*, Nueva York, Chapman-Hall, 1993.
 Kleijnen, J. P. C., *Statistical Techniques in Simulation*, partes 1 y 2, Nueva York, Marcel Dekker, 1974/1975.
 Law, A. M. y W. D. Kelton, *Simulation Modelling and Analysis*, Nueva York, McGraw-Hill, 2a. ed., 1991.

CAPÍTULO

8

Técnicas de reducción de varianza

Introducción

En un caso característico de estudio de simulación, uno está interesado en determinar θ , un parámetro relacionado con algún modelo estocástico. Para estimar θ , el modelo se simula para obtener, entre otras cosas, el dato de salida X tal que $\theta = E[X]$. Se realizan varias ejecuciones de simulación, en las que la i -ésima simulación produce la variable de salida X_i . Luego, el estudio concluye con n ejecuciones y la estimación de θ está dada por $\bar{X} = \Sigma_{i=1}^n X_i / n$. Como esto produce una estimación insesgada de θ , implica que su error cuadrático medio es igual a su varianza. Es decir,

$$\text{ECM} = E[(\bar{X} - \theta)^2] = \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n}$$

Por lo tanto, si podemos obtener otra estimación insesgada de θ con menor varianza que \bar{X} , podríamos obtener un mejor estimador.

En este capítulo presentamos varios métodos que pueden servir para reducir la varianza del estimador de simulación (llamado estimador en bruto) \bar{X} .

Sin embargo, antes de presentar estas técnicas de reducción de varianza, ilustraremos los problemas potenciales, aun en modelos muy sencillos, al utilizar el estimador de simulación en bruto.

Ejemplo 8a Control de calidad. Consideremos un proceso que produce artículos en forma secuencial. Suponga que estos artículos tienen valores medibles asociados y que cuando el proceso está "bajo control" estos valores (normalizados de manera adecuada) provienen de una distribución normal estándar. Suponga además que cuando el proceso está "fuera de control" la distribución de los valores cambia de la normal estándar a alguna otra distribución.

Con frecuencia se utiliza el siguiente procedimiento, la regla de control de promedio móvil con pesos exponenciales, para ver si el proceso se sale de control. Sea X_1, X_2, \dots la sucesión de datos. Para un valor fijo α , $0 \leq \alpha \leq 1$, definimos la sucesión S_n , $n \geq 0$, como

$$S_0 = 0$$

$$S_n = \alpha S_{n-1} + (1 - \alpha)X_n, \quad n \geq 1$$

Cuando el proceso está bajo control, todas las X_n tienen media 0, y así es fácil verificar que, según esta condición, los valores de promedio móvil con pesos exponenciales S_n también tienen media 0. La regla de control de promedio móvil consiste en fijar una constante B , junto con el valor de α , y luego declarar al proceso "fuera de control" cuando $|S_n|$ excede a B . Es decir, el proceso se declara fuera de control en el instante aleatorio N , donde

$$N = \text{Mín}\{n: |S_n| > B\}$$

Es claro que $|S_n|$ será mayor que B en cierto momento, de modo que el proceso será declarado fuera de control aunque esté trabajando en forma adecuada; es decir, incluso cuando los datos sean generados por una distribución normal estándar. Para garantizar que no ocurra con demasiada frecuencia, es prudente elegir α y B de modo que cuando los X_n , $n \geq 1$, realmente provengan de una distribución normal estándar, $E[N]$ sea grande. Suponga que se ha decidido que, en estas condiciones, es aceptable un valor de $E[N] = 800$. Suponga además que se afirma que los valores $\alpha = 0.9$ y $B = 0.8$ producen un valor de $E[N]$ cercano a 800. ¿Cómo podríamos verificar esta afirmación?

Una forma de hacerlo es mediante la simulación. Podemos generar normales estándar X_n , $n \geq 1$, hasta que $|S_n|$ sea mayor que 0.8 (donde $\alpha = 0.9$ en la ecuación que define a S_n). Si N_1 denota el número de normales necesarias hasta que esto ocurre, entonces, para nuestra primera ejecución de simulación, tenemos la variable de salida N_1 . Luego generamos otras ejecuciones y nuestra estimación de $E[N]$ es el valor promedio de los datos de salida obtenidos durante todas las ejecuciones.

Sin embargo, suponga que queremos estar 99 por ciento seguros de que nuestra estimación de $E[N]$, según la hipótesis de que el sistema está bajo control, es precisa salvo un error de ± 0.1 . Por lo tanto, como 99 por ciento del tiempo una variable aleatoria normal está ± 2.33 desviaciones estándares de su media (es decir, $z_{0.005} = 2.33$), el número de ejecuciones necesarias (denotado n) es tal que

$$\frac{2.33\sigma_n}{\sqrt{n}} \approx 0.1$$

donde σ_n es la desviación estándar muestral basada en los primeros n datos. Ahora, σ_n será aproximadamente igual a $\sigma(N)$, la desviación estándar de N , y ahora es posible argumentar que esto es aproximadamente igual a $E[N]$. El argumento es el siguiente: como estamos suponiendo que el proceso permanece bajo control todo el tiempo, la mayor parte del tiempo ocurre que el valor del promedio móvil con pesos exponenciales está cerca del origen. En ciertas ocasiones, de manera aleatoria, crece y tiende, en valor absoluto, a B . En tales ocasiones, puede exceder B , en cuyo caso la ejecución termina; o bien, podría haber una cadena de datos normales tales que, después de un breve lapso, eliminan el hecho de que el promedio móvil ha sido grande (esto se debe a que los valores anteriores de S_j se multiplican de manera continua por 0.9 y por tanto pierden su efecto). Por lo tanto, si sabemos que el proceso no se ha salido aún de control durante cierto tiempo fijo k , entonces, sin importar el valor de k , parecería que el valor de S_k está cerca del origen. En otras palabras, es intuitivamente claro que la distribución del tiempo hasta que el promedio móvil excede los límites de control no tiene memoria; es decir, es aproximadamente una variable aleatoria exponencial. Pero para una variable aleatoria exponencial Y , $\text{Var}(Y) = (E[Y])^2$. Como la desviación estándar es la raíz cuadrada de la varianza, parece intuitivamente claro que, si el sistema está bajo control todo el tiempo, $\sigma(N) \approx E[N]$. Por lo tanto, si la afirmación original de que $E[N] \approx 800$ es correcta, el número de ejecuciones necesarias es tal que

$$\sqrt{n} \approx 23.3 \times 800$$

o

$$n \approx (23.3 \times 800)^2 \approx 3.47 \times 10^8$$

Además, como cada ejecución requiere aproximadamente 800 variables aleatorias normales (de nuevo, suponiendo que la afirmación es casi correcta), vemos que para hacer esta simulación se necesitan aproximadamente $800 \times 3.47 \times 10^8 \approx 2.77 \times 10^{11}$ variables aleatorias normales, una tarea gigantesca. ■

8.1 El uso de variables antitéticas

Suponga que estamos interesados en utilizar la simulación para estimar $\theta = E[X]$ y suponga que hemos generado X_1 y X_2 , variables aleatorias idénticamente distribuidas con media θ . Entonces

$$\text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4} [\text{Var}(X_1) + \text{Var}(X_2) + 2 \text{Cov}(X_1, X_2)]$$

Por lo tanto, sería bueno (en el sentido de que la varianza se reduce) que X_1 y X_2 estuvieran correlacionadas en forma negativa en vez de ser independientes.

Para lograr que X_1 y X_2 estén correlacionadas en forma negativa, suponga que X_1 es una función de m números aleatorios; es decir, suponga que

$$X_1 = h(U_1, U_2, \dots, U_m)$$

donde U_1, \dots, U_m son m números aleatorios arbitrarios. Ahora, si U es un número aleatorio —es decir, U está uniformemente distribuido en $(0, 1)$ —, entonces también lo está $1 - U$. Por lo tanto, la variable aleatoria

$$X_2 = h(1 - U_1, 1 - U_2, \dots, 1 - U_m)$$

tiene la misma distribución que X_1 . Además, como claramente $1 - U$ está correlacionada de manera negativa con U , esperaríamos que X_2 estuviese correlacionada de manera negativa con X_1 ; de hecho, este resultado se puede demostrar en el caso particular de que h sea una función monótona (creciente o decreciente) de cada una de sus coordenadas. [Este resultado es consecuencia de un resultado más general, el cual establece que dos funciones crecientes (o decrecientes) de un conjunto de variables aleatorias independientes están correlacionadas en forma positiva. Ambos resultados aparecen en el apéndice.] Por lo tanto, en este caso, después de generar U_1, \dots, U_m para calcular X_1 , en vez de generar un nuevo conjunto independiente de m números aleatorios, podemos hacerlo utilizando solamente el conjunto $1 - U_1, \dots, 1 - U_m$ para calcular a X_2 . Además, debemos observar que conseguimos un beneficio doble; no sólo obtenemos un estimador con menor varianza (al menos cuando h es una función monótona), sino que ahorramos el tiempo necesario para generar otro conjunto de números aleatorios.

Ejemplo 8b Simulación de la función de fiabilidad. Consideremos un sistema de n componentes, cada uno de los cuales está funcionando o está descompuesto. Si

$$s_i = \begin{cases} 1 & \text{si el componente } i \text{ funciona} \\ 0 & \text{en caso contrario} \end{cases}$$

decimos que $s = (s_1, \dots, s_n)$ es el vector de estado. Suponga además que existe una función no decreciente $\phi(s_1, \dots, s_n)$ tal que

$$\phi(s_1, \dots, s_n) = \begin{cases} 1 & \text{si el sistema funciona bajo el vector de estado } s_1, \dots, s_n \\ 0 & \text{en caso contrario} \end{cases}$$

La función $\phi(s_1, \dots, s_n)$ es la función de estructura.

Algunas funciones de estructura comunes son las siguientes:

(a) *La estructura en serie.* Para la estructura en serie

$$\phi(s_1, \dots, s_n) = \text{Mín}_i s_i$$

El sistema en serie sólo funciona si todos sus componentes funcionan.

(b) *La estructura en paralelo.* Para la estructura en paralelo,

$$\phi(s_1, \dots, s_n) = \text{Máx}_i s_i$$

Por lo tanto, el sistema en paralelo funciona si al menos uno de sus componentes funciona.

(c) *El sistema k de n .* La función de estructura

$$\phi(s_1, \dots, s_n) = \begin{cases} 1 & \text{Si } \sum_{i=1}^n s_i \geq k \\ 0 & \text{en caso contrario} \end{cases}$$

es una función de estructura k de n . Como $\sum_{i=1}^n s_i$ representa la cantidad de componentes funcionando, un sistema k de n funciona si al menos k de los n componentes están funcionando.

Hay que observar que un sistema en serie es un sistema n de n , mientras que un sistema en paralelo es un sistema 1 de n .

(d) *La estructura de puente.* Un sistema de cinco componentes para el cual

$$\phi(s_1, s_2, s_3, s_4, s_5) = \text{Máx}(s_1 s_3 s_5, s_2 s_3 s_4, s_1 s_4, s_2 s_5)$$

tiene una estructura de puente. Tal sistema se puede representar en forma esquemática como en la figura 8.1. La idea del diagrama es que el sistema funciona si una señal puede ir de izquierda a derecha de éste. La señal puede pasar por cualquier nodo i , siempre que el componente i esté funcionando. Dejaremos como ejercicio para el lector verificar la fórmula dada para la función de estructura de puente.

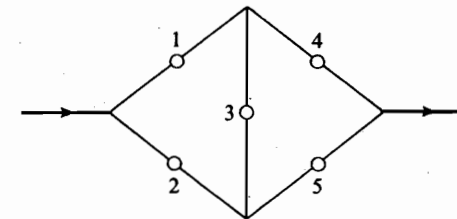


Figura 8.1 La estructura de puente

Ahora suponga que los estados de los componentes (denotados S_i) $i = 1, \dots, n$, son variables aleatorias independientes tales que

$$P\{S_i = 1\} = p_i = 1 - P\{S_i = 0\} \quad i = 1, \dots, n$$

Sea

$$\begin{aligned} r(p_1, \dots, p_n) &= P\{\phi(S_1, \dots, S_n) = 1\} \\ &= E[\phi(S_1, \dots, S_n)] \end{aligned}$$

La función $r(p_1, \dots, p_n)$ es la función de *fiabilidad*. Representa la probabilidad de que el sistema funcione cuando los componentes son independientes, donde el componente i funciona con probabilidad p_i , $i = 1, \dots, n$.

Para un sistema en serie,

$$\begin{aligned} r(p_1, \dots, p_n) &= P\{S_i = 1 \text{ para toda } i = 1, \dots, n\} \\ &= \prod_{i=1}^n P\{S_i = 1\} \\ &= \prod_{i=1}^n p_i \end{aligned}$$

y para un sistema en paralelo,

$$\begin{aligned} r(p_1, \dots, p_n) &= P\{S_i = 1 \text{ para al menos una } i = 1, \dots, n\} \\ &= 1 - P\{S_i = 0 \text{ para toda } i = 1, \dots, n\} \\ &= 1 - \prod_{i=1}^n P\{S_i = 0\} \\ &= 1 - \prod_{i=1}^n (1 - p_i) \end{aligned}$$

Sin embargo, para la mayor parte de los sistemas es un enorme problema el cálculo de la función de fiabilidad (incluso para sistemas pequeños como el sistema cinco de 10 o el sistema de puente podría ser tedioso su cálculo). Así, suponga que para una función de estructura ϕ dada, no decreciente, y las probabilidades dadas p_1, \dots, p_n , nos interesa utilizar la simulación para estimar

$$r(p_1, \dots, p_n) = E[\phi(S_1, \dots, S_n)]$$

Para estimar los S_i generamos números aleatorios uniformes U_1, \dots, U_n y hacemos

$$S_i = \begin{cases} 1 & \text{si } U_i < p_i \\ 0 & \text{en caso contrario} \end{cases}$$

Por lo tanto, vemos que

$$\phi(S_1, \dots, S_n) = h(U_1, \dots, U_n)$$

donde h es una función decreciente de U_1, \dots, U_n . Por lo tanto,

$$\text{Cov}(h(U), h(1 - U)) \leq 0$$

y así, el método de las variables antitéticas consistente en utilizar U_1, \dots, U_n para generar $h(U_1, \dots, U_n)$ y $h(1 - U_1, \dots, 1 - U_n)$ produce una menor varianza que en el caso en el que empleamos un conjunto independiente de números aleatorios para generar el segundo valor de h . ■

Con frecuencia, la salida importante de una simulación es una función de las variables aleatorias de entrada Y_1, \dots, Y_m ; es decir, la salida importante es $X = h(Y_1, \dots, Y_m)$. Suponga que las Y_i tienen distribución F_i , $i = 1, \dots, m$. Si estos valores de entrada se generan mediante la técnica de la transformada inversa, podemos escribir

$$X = h(F_1^{-1}(U_1), \dots, F_m^{-1}(U_m))$$

donde U_1, \dots, U_m son números aleatorios independientes. Como una función de distribución es creciente, su inversa también es creciente, de modo que si $h(y_1, \dots, y_m)$ fuese una función monótona de sus coordenadas, se tendría que $h(F_1^{-1}(U_1), \dots, F_m^{-1}(U_m))$ sería una función monótona de las U_i . Por lo tanto, el método de las variables antitéticas, el cual generaría primero U_1, \dots, U_m para calcular X_1 y luego tomaría $1 - U_1, \dots, 1 - U_m$ para calcular X_2 , daría como resultado un estimador con una menor varianza que el obtenido si se usa un nuevo conjunto de números aleatorios para X_2 .

Ejemplo 8c Simulación de un sistema de línea de espera. Consideremos un sistema de línea de espera dado, y sea D_i el retraso, tiempo que permanece esperando en la cola, del i -ésimo cliente, y suponga que estamos interesados en simular el sistema para estimar $\theta = E[X]$, donde

$$X = D_1 + \dots + D_n$$

es la suma de los retrasos en la cola de las n primeras llegadas. Sean I_1, \dots, I_n los primeros n tiempos entre las llegadas (es decir, I_j es el tiempo entre las llegadas de los clientes $j - 1$ y j) y sean S_1, \dots, S_n los n primeros tiempos de servicio de este sistema;

suponga que todas estas variables aleatorias son independientes. En muchos sistemas, X es una función de las $2n$ variables aleatorias $I_1, \dots, I_n, S_1, \dots, S_n$, digamos,

$$X = h(I_1, \dots, I_n, S_1, \dots, S_n)$$

Además, como por lo general el retraso de un cliente dado en la cola aumenta (por supuesto, según los datos específicos del modelo) cuando los tiempos de servicio de otros clientes aumentan y por lo general decrece cuando los tiempos entre las llegadas aumentan, esto implica que, para muchos modelos, h es una función monótona de sus coordenadas. Por lo tanto, si se aplica el método de la transformada inversa para generar las variables aleatorias $I_1, \dots, I_n, S_1, \dots, S_n$, entonces el método de las variables antitéticas produce una menor varianza. Es decir, si inicialmente tomamos los $2n$ números aleatorios $U_i, i = 1, \dots, 2n$, para generar los tiempos entre llegadas y los tiempos de servicio haciendo $I_i = F_i^{-1}(U_i)$, $S_i = G_i^{-1}(U_{n+i})$, donde F_i y G_i son respectivamente las funciones de distribución de I_i y S_i , entonces la segunda ejecución de simulación debe realizarse de la misma manera, pero con los números aleatorios $1 - U_i, i = 1, \dots, 2n$. Esto produce una menor varianza que si se genera un nuevo conjunto de $2n$ números aleatorios para la segunda ejecución. ■

El siguiente ejemplo ilustra la mejora que se logra en ciertas ocasiones con el uso de las variables antitéticas.

Ejemplo 8d Suponga que estamos interesados en realizar una simulación para estimar

$$\theta = E[e^U] = \int_0^1 e^x dx$$

(Por supuesto, sabemos que $\theta = e - 1$; sin embargo, el punto de este ejemplo es ver la mejora posible al utilizar las variables antitéticas.) Como es claro que la función $h(u) = e^u$ es una función monótona, el método de las variables antitéticas conduce a una reducción de varianza, cuyo valor determinamos a continuación. Para comenzar, observemos que

$$\begin{aligned} \text{Cov}(e^U, e^{1-U}) &= E[e^U e^{1-U}] - E[e^U]E[e^{1-U}] \\ &= e - (e - 1)^2 = -0.2342 \end{aligned}$$

Además, como

$$\begin{aligned} \text{Var}(e^U) &= E[e^{2U}] - (E[e^U])^2 \\ &= \int_0^1 e^{2x} dx - (e - 1)^2 \end{aligned}$$

$$= \frac{[e^2 - 1]}{2} - (e - 1)^2 = 0.2420$$

vemos que el uso de números aleatorios independientes produce una varianza de

$$\text{Var}\left(\frac{\exp\{U_1\} + \exp\{U_2\}}{2}\right) = \frac{\text{Var}(e^U)}{2} = 0.1210$$

mientras que el uso de las variables antitéticas U y $1 - U$ produce una varianza de

$$\text{Var}\left(\frac{e^U + e^{1-U}}{2}\right) = \frac{\text{Var}(e^U)}{2} + \frac{\text{Cov}(e^U, e^{1-U})}{2} = 0.0039$$

con una reducción de varianza de 96.7 por ciento. ■

Ejemplo 8e Estimación de e . Consideremos una sucesión de números aleatorios y sea N el primero que es mayor que su predecesor inmediato. Es decir,

$$N = \min(n : n \geq 2, U_n > U_{n-1})$$

Ahora,

$$\begin{aligned} P\{N > n\} &= P\{U_1 \geq U_2 \geq \dots \geq U_n\} \\ &= 1/n! \end{aligned}$$

donde la última igualdad se cumple pues todos los posibles ordenamientos de U_1, \dots, U_n son igualmente probables. Por lo tanto,

$$P\{N = n\} = P\{N > n - 1\} - P\{N > n\} = \frac{1}{(n - 1)!} - \frac{1}{n!} = \frac{n - 1}{n!}$$

y entonces

$$E[N] = \sum_{n=2}^{\infty} \frac{1}{(n - 2)!} = e$$

Además,

$$\begin{aligned} E[N^2] &= \sum_{n=2}^{\infty} \frac{n}{(n - 2)!} = \sum_{n=2}^{\infty} \frac{2}{(n - 2)!} + \sum_{n=2}^{\infty} \frac{n - 2}{(n - 2)!} \\ &= 2e + \sum_{n=3}^{\infty} \frac{1}{(n - 3)!} = 3e \end{aligned}$$

y entonces

$$\text{Var}(N) = 3e - e^2 \approx 0.7658$$

Por lo tanto, e se puede estimar al generar números aleatorios y detenerse la primera vez que uno es mayor que su predecesor inmediato.

Si empleamos variables antitéticas, entonces también podríamos hacer

$$M = \min(n: n \geq 2, 1 - U_n > 1 - U_{n-1}) = \min(n: n \geq 2, U_n < U_{n-1})$$

Como uno de los valores N y M será igual a 2 y el otro será mayor que 2, parecería, aunque no son funciones monótonas de las U_n , que el estimador $(N + M)/2$ tendría una menor varianza que el promedio de las dos variables aleatorias independientes distribuidas de acuerdo con N . Antes de determinar $\text{Var}(N + M)$, es útil considerar primero la variable aleatoria N_a , cuya distribución es igual a la distribución condicional de la cantidad de números aleatorios adicionales que deben observarse hasta que uno de los números observados sea mayor que su predecesor, dado que $U_2 \leq U_1$. Por lo tanto, podemos escribir

$$N = 2, \quad \text{con probabilidad } \frac{1}{2}$$

$$N = 2 + N_a, \quad \text{con probabilidad } \frac{1}{2}$$

Por lo tanto,

$$E[N] = 2 + \frac{1}{2} E[N_a]$$

$$E[N^2] = \frac{1}{2} 4 + \frac{1}{2} E[(2 + N_a)^2]$$

$$= 4 + 2E[N_a] + \frac{1}{2} E[N_a^2]$$

Con los resultados obtenidos para $E[N]$ y $\text{Var}(N)$ obtenemos, después de un poco de álgebra,

$$E[N_a] = 2e - 4$$

$$E[N_a^2] = 8 - 2e$$

lo cual implica que

$$\text{Var}(N_a) = 14e - 4e^2 - 8 \approx 0.4997$$

Ahora consideremos las variables aleatorias N y M . Es fácil ver que después de observar los dos primeros números aleatorios, una de las variables N y M será igual a 2 y la otra será igual a 2 más una variable aleatoria que tiene la misma distribución que N_a . Por lo tanto,

$$\text{Var}(N + M) = \text{Var}(4 + N_a) = \text{Var}(N_a)$$

Por lo tanto,

$$\frac{\text{Var}(N_1 + N_2)}{\text{Var}(N + M)} \approx \frac{1.5316}{0.4997} \approx 3.065$$

Así, el uso de variables antitéticas reduce la varianza del estimador por un factor ligeramente mayor que 3. ■

En el caso de una variable aleatoria normal con media μ y varianza σ^2 , podemos servirnos del método de las variables antitéticas: generamos primero tal variable aleatoria Y y luego consideramos como variable antitética $2\mu - Y$, la cual también es normal con media μ y varianza σ^2 y que está correlacionada claramente en forma negativa con Y . Si utilizamos la simulación para calcular $E[h(Y_1, \dots, Y_n)]$, donde las Y_i son variables aleatorias normales independientes y h es una función monótona de sus coordenadas, entonces el método antitético consistente en generar primero las n normales Y_1, \dots, Y_n para calcular $h(Y_1, \dots, Y_n)$ y luego utilizar las variables antitéticas $2\mu_i - Y_i, i = 1, \dots, n$, para calcular el siguiente valor simulado de h conduciría a una reducción de varianza en comparación con la generación de un segundo conjunto de n variables aleatorias normales.

8.2 El uso de variables de control

De nuevo, suponga que queremos utilizar simulación para estimar $\theta = E[X]$, donde X es la salida de la simulación. Suponga que se conoce el valor esperado de alguna otra variable de salida Y , digamos, $E[Y] = \mu_y$. Entonces, para cualquier constante c , la cantidad

$$X + c(Y - \mu_y)$$

también es un estimador insesgado de θ . Para determinar el mejor valor de c , observe que

$$\begin{aligned} \text{Var}(X + c(Y - \mu_y)) &= \text{Var}(X + cY) \\ &= \text{Var}(X) + c^2 \text{Var}(Y) + 2c \text{Cov}(X, Y) \end{aligned}$$

Una aplicación sencilla del cálculo nos muestra que lo anterior se minimiza cuando $c = c^*$, donde

$$c^* = - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \quad (8.1)$$

y para este valor la varianza del estimador es

$$\text{Var}(X + c^*(Y - \mu_y)) = \text{Var}(X) - \frac{[\text{Cov}(X, Y)]^2}{\text{Var}(Y)} \quad (8.2)$$

La cantidad Y es una *variable de control* para el estimador de simulación X . Para ver por qué funciona, observe que c^* es negativa (positiva) cuando X y Y están correlacionadas en forma positiva (negativa). Así, suponga que X y Y están correlacionadas en forma positiva, lo cual significa, en términos generales, que X es grande cuando Y es grande y viceversa. Por lo tanto, si una ejecución de simulación produce un valor grande (pequeño) de Y , lo cual queda indicado por el hecho de que Y sea mayor que su media conocida μ_y , entonces es probable que X también sea mayor (menor) que su media θ ; quisiéramos corregir esto reduciendo (aumentando) el valor del estimador X , lo cual se logra pues c^* es negativa (positiva). Se tiene un argumento similar cuando X y Y están correlacionadas en forma negativa.

Al dividir la ecuación (8.2) entre $\text{Var}(X)$, obtenemos

$$\frac{\text{Var}(X + c^*(Y - \mu_y))}{\text{Var}(X)} = 1 - \text{Corr}^2(X, Y)$$

donde

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

es la correlación entre X y Y . Por lo tanto, la reducción de varianza obtenida al utilizar la variable de control Y es de $100 \text{ Corr}^2(X, Y)$ por ciento.

Por lo general, las cantidades $\text{Cov}(X, Y)$ y $\text{Var}(X, Y)$ no se conocen de antemano y deben estimarse a partir de los datos simulados. Si se realizan n ejecuciones de simulación y se obtienen los datos de salida $X_i, Y_i, i = 1, \dots, n$, entonces podemos utilizar los estimadores

$$\widehat{\text{Cov}}(X, Y) = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / (n - 1)$$

y

$$\widehat{\text{Var}}(Y) = \sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1),$$

para aproximar c^* mediante \hat{c}^* , donde

$$\hat{c}^* = - \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}.$$

La varianza del estimador controlado

$$\text{Var}(\bar{X} + c^*(\bar{Y} - \mu_y)) = \frac{1}{n} \left(\text{Var}(X) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(Y)} \right)$$

se puede estimar mediante el estimador de $\text{Cov}(X, Y)$ junto con los estimadores de la varianza muestral de $\text{Var}(X)$ y $\text{Var}(Y)$.

Observación Otra forma de hacer los cálculos es utilizar un paquete estándar de computadora para modelos sencillos de regresión lineal, ya que si consideramos el modelo de regresión lineal simple

$$X = a + bY + e$$

donde e es una variable aleatoria con media 0 y varianza σ^2 , entonces \hat{a} y \hat{b} , los estimadores por mínimos cuadrados de a y b con base en los datos $X_i, Y_i, i = 1, \dots, n$, son

$$\hat{b} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

$$\hat{a} = \bar{X} - \hat{b}\bar{Y}$$

Por lo tanto, $\hat{b} = -\hat{c}^*$. Además, como

$$\begin{aligned} \bar{X} + \hat{c}^*(\bar{Y} - \mu_y) &= \bar{X} - \hat{b}(\bar{Y} - \mu_y) \\ &= \hat{a} + \hat{b}\mu_y \end{aligned}$$

se tiene que la estimación de la variable de control es la evaluación de la recta de regresión estimada en el valor $Y = \mu_y$. Además, como $\hat{\sigma}^2$, la estimación por regresión de σ^2 , es la estimación de $\text{Var}(X - \hat{b}Y) = \text{Var}(X + \hat{c}^*Y)$, se tiene que la varianza estimada del estimador de la variable de control $\bar{X} + \hat{c}^*(\bar{Y} - \mu_y)$ es $\hat{\sigma}^2/n$. ■

Ejemplo 8f Suponga, como en el ejemplo 8b, que queremos realizar una simulación para estimar la función de fiabilidad

$$r(p_1, \dots, p_n) = E[\phi(S_1, \dots, S_n)]$$

donde

$$S_i = \begin{cases} 1 & \text{si } U_i < p_i \\ 0 & \text{en caso contrario} \end{cases}$$

Como $E[S_i] = p_i$, se tiene que

$$E\left[\sum_{i=1}^n S_i\right] = \sum_{i=1}^n p_i$$

Por lo tanto, podemos utilizar el número de componentes que funcionan, $Y \equiv \sum S_i$, como variable de control del estimador $X \equiv \phi(S_1, \dots, S_n)$. Como $\sum_{i=1}^n S_i$ y $\phi(S_1, \dots, S_n)$ son ambas funciones crecientes de las S_i , están correlacionadas en forma positiva, y así el signo de c^* es negativo. ■

Ejemplo 8g Consideremos un sistema de línea de espera en la cual los clientes llegan de acuerdo con un proceso Poisson no homogéneo con función de intensidad $\lambda(s)$, $s > 0$. Suponga que los tiempos de servicio son variables aleatorias independientes con distribución G y que también son independientes de los tiempos de llegada. Suponga que estamos interesados en estimar el tiempo total que pasan dentro del sistema todos los clientes que lleguen después del tiempo t . Es decir, si W_i denota la cantidad de tiempo que el i -ésimo cliente pasa dentro del sistema, entonces estamos interesados en $\theta = E[X]$, donde

$$X = \sum_{i=1}^{N(t)} W_i$$

y donde $N(t)$ es el número de llegadas hasta el tiempo t . Una cantidad natural para utilizar como control en esta situación es el total de los tiempos de servicio de todos estos clientes. Es decir, sea S_i el tiempo de servicio del i -ésimo cliente y sea

$$Y = \sum_{i=1}^{N(t)} S_i$$

Como los tiempos de servicio son independientes de $N(t)$, se tiene que

$$E[Y] = E[S]E[N(t)]$$

donde $E[S]$, el tiempo medio de servicio, y $E[N(t)]$, el número promedio de llegadas hasta t , son ambas cantidades conocidas. ■

Ejemplo 8h Como en el ejemplo 8d, suponga que estamos interesados en ejecutar una simulación para calcular $\theta = E[e^U]$. En este caso, una variable natural para utilizar

como control es el número aleatorio U . Para apreciar la mejora posible con respecto del estimador en bruto, observe que

$$\begin{aligned} \text{Cov}(e^U, U) &= E[Ue^U] - E[U]E[e^U] \\ &= \int_0^1 xe^x dx - \frac{(e-1)}{2} \\ &= 1 - \frac{(e-1)}{2} = 0.14086 \end{aligned}$$

Como $\text{Var}(U) = \frac{1}{12}$, (8.2) implica que

$$\begin{aligned} \text{Var}\left(e^U + c^*\left(U - \frac{1}{2}\right)\right) &= \text{Var}(e^U) - 12(0.14086)^2 \\ &= 0.2420 - 0.2380 = 0.0039 \end{aligned}$$

donde hemos tomado $\text{Var}(e^U) = 0.2420$ (ejemplo 8d). Por lo tanto, en este caso, el uso de la variable de control U puede conducir a una reducción de la varianza de hasta 98.4 por ciento. ■

Ejemplo 8i Un problema de reordenamiento de listas. Suponga dado un conjunto de n elementos, numerados de 1 a n , que deben formar una lista ordenada. En cada unidad de tiempo se realiza la solicitud de recuperación de uno de estos elementos, donde la solicitud del elemento i tiene la probabilidad $p(i)$, $\sum_{i=1}^n p(i) = 1$. Después de ser solicitado, el elemento regresa a la lista, pero no necesariamente en la misma posición. Por ejemplo, una regla común de reordenamiento es intercambiar el elemento solicitado con el inmediato anterior. Así, si $n = 4$ y el ordenamiento actual es 1, 4, 2, 3, entonces esta regla implica que una solicitud del elemento 2 produce el reordenamiento 1, 2, 4, 3. Si se parte de un ordenamiento inicial, donde cualquiera de los $n!$ ordenamientos tiene la misma probabilidad, y se sigue esta regla de intercambio, suponga que estamos interesados en determinar la suma esperada de las posiciones de los primeros N elementos solicitados. ¿Cómo podemos realizar esto de manera eficiente mediante simulación?

Una forma eficiente es la que sigue. La forma "natural" de simular lo anterior es generar primero una permutación aleatoria de 1, 2, ..., n para establecer el ordenamiento inicial y luego, en cada uno de los siguientes N periodos, para determinar el elemento solicitado se genera un número aleatorio U y se solicita el elemento j si $\sum_{k=1}^{j-1} p(k) < U \leq \sum_{k=1}^j p(k)$. Sin embargo, una técnica mejor consiste en generar el elemento solicitado de modo que los valores pequeños de U correspondan a elementos cercanos al frente. Específicamente, si el ordenamiento actual es i_1, i_2, \dots, i_n , entonces para producir el elemento solicitado generamos un número aleatorio U y solicitamos i_j si $\sum_{k=1}^{j-1} p(i_k) < U \leq \sum_{k=1}^j p(i_k)$. Por ejemplo, si $n = 4$ y el ordenamiento

actual es 3, 1, 2, 4, entonces deberíamos generar U y seleccionar 3 si $U \leq p(3)$, seleccionar 1 si $p(3) < U \leq p(3) + p(1)$, y así sucesivamente. Como de esta forma los valores pequeños de U corresponden a elementos cerca del frente, podemos emplear $\sum_{r=1}^N U_r$ como variables de control, donde U_r es el número aleatorio utilizado para la r -ésima solicitud en una ejecución. Es decir, si P_r es la posición del r -ésimo elemento elegido en una ejecución, entonces en vez de utilizar el estimador en bruto $\sum_{r=1}^N P_r$, deberíamos utilizar

$$\sum_{r=1}^N P_r + c^* \left(\sum_{r=1}^N U_r - \frac{N}{2} \right)$$

donde

$$c^* = - \frac{\text{Cov} \left(\sum_{r=1}^N P_r, \sum_{r=1}^N U_r \right)}{\frac{N}{12}}$$

y donde la varianza anterior se debe estimar mediante los datos de todas las ejecuciones simuladas.

Aunque la reducción de varianza obtenida dependerá, por supuesto, de las probabilidades $p(i)$, $i = 1, \dots, n$, y del valor de N , un estudio somero indica que cuando $n = 50$ y las $p(i)$ son aproximadamente iguales, entonces para $15 \leq N \leq 50$ la varianza del estimador controlado es menor que $\frac{1}{2400}$ de la varianza del estimador de simulación en bruto. ■

Por supuesto, uno puede utilizar más de una variable como control. Por ejemplo, si una simulación produce las variables de salida Y_i , $i = 1, \dots, k$, y si se conoce $E[Y_i] = \mu_i$, entonces para cualesquiera constantes c_i , $i = 1, \dots, k$, nos valemos de

$$X + \sum_{i=1}^k c_i (Y_i - \mu_i)$$

como estimador insesgado de $E[X]$.

Ejemplo 8j Blackjack. Con frecuencia, el juego de blackjack se juega de este modo: se barajan varios conjuntos de naipes, se apartan las cartas utilizadas y se vuelve a barajar cuando el número de cartas restantes está por debajo de cierto límite; digamos también que inicia una nueva ronda cada vez que el tallador vuelve a barajar las cartas. Ahora, suponga que estamos interesados en realizar una simulación para estimar $E[X]$, las ganancias esperadas de un jugador por cada ronda, y suponemos que practica cierta estrategia fija, la cual podría ser del tipo "contar las cartas" que ya han

aparecido en la ronda y apostar diversas cantidades según la "cuenta". Supondremos que el juego consta de un único jugador contra el tallador.

Lo aleatorio del juego se debe al barajado de las cartas por parte del tallador. Si éste utiliza k barajas de 52 cartas cada una, entonces podemos generar el barajado mediante una permutación aleatoria de los números 1 a $52k$; sea I_1, \dots, I_{52k} esta permutación. Si ahora hacemos

$$u_j = I_j \bmod 13 + 1$$

y

$$v_j = \min(u_j, 10)$$

entonces v_j , $j = 1, \dots, 52k$ representa los valores sucesivos de las cartas barajadas, donde 1 equivale a un as.

Sea N el número de manos jugadas en una ronda y B_j la cantidad apostada en la mano j . Para reducir la varianza, usamos una variable de control que sea grande cuando el jugador tiene una mejor mano que el tallador y que sea pequeña en el caso contrario. Como tener 19 puntos o más es bueno, definimos

$$W_j = 1 \text{ si las dos cartas del jugador en su turno } j \text{ suman al menos } 19$$

y sea $W_j = 0$ en caso contrario. De manera similar, sea

$$Z_j = 1 \text{ si las dos cartas del tallador en su turno } j \text{ suman al menos } 19$$

y sea $Z_j = 0$ en caso contrario. Como es claro que W_j y Z_j tienen la misma distribución, esto implica que $E[W_j - Z_j] = 0$, y no es difícil mostrar que

$$E \left[\sum_{j=1}^N B_j (W_j - Z_j) \right] = 0$$

Así, recomendamos el uso de $\sum_{j=1}^N B_j (W_j - Z_j)$ como variable de control. Por supuesto, no es claro que 19 sea el mejor valor, y uno debería experimentar cuando 18 o incluso 20 sean el valor crítico. Sin embargo, un poco de trabajo preliminar indica que 19 funciona mejor y produce reducciones de varianza de 15 por ciento o más según la estrategia utilizada por el jugador. Se podría obtener una reducción de varianza mucho mayor si utilizamos dos variables de control. Una variable de control se define como antes, con la excepción de que los W_j y Z_j se definen como 1 si la mano es 19 o 20. La segunda variable de control es similar, pero esta vez los indicadores valen 1 si las manos constan de blackjacks. ■

Cuando se utilizan varias variables de control, los cálculos se realizan mediante un programa de computadora para el modelo de regresión lineal múltiple

$$X = a + \sum_{i=1}^k b_i Y_i + e$$

donde e es una variable aleatoria con media 0 y varianza σ^2 . Si \hat{c}_i^* es la estimación de la mejor c_i para $i = 1, \dots, k$, entonces

$$\hat{c}_i^* = -\hat{b}_i, \quad i = 1, \dots, k$$

donde \hat{b}_i , $i = 1, \dots, k$, son las estimaciones de b_i , $i = 1, \dots, k$ obtenidas mediante regresión por mínimos cuadrados. El valor de la estimación controlada se puede obtener a partir de

$$\bar{X} + \sum_{i=1}^k \hat{c}_i^* (\bar{Y}_i - \mu_i) = \hat{a} + \sum_{i=1}^k \hat{b}_i \mu_i$$

Es decir, la estimación controlada no es más que la recta de regresión múltiple evaluada en el punto (μ_1, \dots, μ_k) .

La varianza de la estimación controlada se puede obtener al dividir la estimación por regresión de σ^2 entre el número de ejecuciones de simulación.

Observaciones

1. Como la varianza de la estimación controlada no se conoce de antemano, con frecuencia la simulación se realiza en dos etapas. En la primera etapa se realiza un pequeño número de ejecuciones para obtener una estimación preliminar de $\text{Var}(X + c^*(Y - \mu_y))$ (esta estimación se obtiene mediante un programa de regresión lineal simple, donde Y es la variable independiente y X es la variable dependiente, utilizando la estimación de σ^2). Luego fijamos el número de ensayos necesarios en la segunda ejecución, de modo que la varianza del estimador final esté dentro de un límite aceptable.
2. Una forma valiosa de interpretar el método de las variables de control es que combina estimadores de θ . Es decir, suponga que los valores de X y W quedan determinados mediante la simulación, y que $E[X] = E[W] = \theta$. Entonces podemos considerar cualquier estimador insesgado de la forma

$$\alpha X + (1 - \alpha)W$$

El mejor de tales estimadores, obtenido al elegir el valor de α que minimice la varianza, está dado por $\alpha = \alpha^*$, donde

$$\alpha^* = \frac{\text{Var}(W) - \text{Cov}(X, W)}{\text{Var}(X) + \text{Var}(W) - 2 \text{Cov}(X, W)} \quad (8.3)$$

Ahora, si se conoce $E[Y] = \mu_y$, tenemos los dos estimadores insesgados X y $X + Y - \mu_y$. El estimador combinado se puede escribir entonces como

$$(1 - c)X + c(X + Y - \mu_y) = X + c(Y - \mu_y)$$

Para ir en la otra dirección de la equivalencia entre las variables de control y la combinación de los estimadores, suponga que $E[X] = E[W] = \theta$. Entonces, si utilizamos X y controlamos mediante la variable $Y = X - W$, la cual sabemos que tiene media 0, obtenemos un estimador de la forma

$$X + c(X - W) = (1 + c)X - cW$$

que es un estimador combinado con $\alpha = 1 + c$.

3. Con la interpretación dada en la observación 2, el método de las variables antitéticas se puede considerar como un caso particular de las variables de control. Es decir, si $E[X] = \theta$, donde $X = h(U_1, \dots, U_n)$, entonces también $E[W] = \theta$, donde $W = h(1 - U_1, \dots, 1 - U_n)$. Por lo tanto, es posible combinar para obtener un estimador de la forma $\alpha X + (1 - \alpha)W$. Como $\text{Var}(X) = \text{Var}(W)$, cuando X y W tienen la misma distribución, la ecuación (8.3) implica que el mejor valor de α es $\alpha = \frac{1}{2}$, y éste es el estimador para las variables antitéticas. ■

8.3 Reducción de varianza mediante condicionamiento

Recordemos la fórmula de la varianza condicional demostrada en la sección 2.10 del capítulo 2.

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

Como los dos términos de la derecha son no negativos, pues una varianza siempre es no negativa, vemos que

$$\text{Var}(X) \geq \text{Var}(E[X|Y]) \quad (8.4)$$

Ahora, suponga que estamos interesados en realizar un estudio de simulación para verificar el valor de $\theta = E[X]$, donde X es una variable de salida de una ejecución de simulación. Además, suponga que hay otra variable Y , tal que $E[X|Y]$ es conocida y toma un valor que se puede determinar a partir de la ejecución de simulación. Como

$$E[E[X|Y]] = E[X] = \theta$$

esto implica que $E[X|Y]$ es también un estimador insesgado de θ ; así, (8.4) implica que como estimador de θ , $E[X|Y]$ es superior al estimador (en bruto) X .

Observaciones Para entender por qué el estimador de la esperanza condicional es superior al estimador en bruto, observe primero que estamos realizando la simulación para estimar el valor desconocido de $E[X]$. Ahora, imaginemos que una ejecución de simulación se realiza en dos etapas: primero observamos el valor simulado de la variable aleatoria Y y luego el valor simulado de X . Sin embargo, después de observar Y , si podemos calcular el valor esperado (condicional) de X , entonces lo aprovechamos para obtener una estimación de $E[X]$, lo cual elimina la varianza adicional contenida en la simulación del valor real de X . ■

En este momento podríamos intentar otras mejoras mediante un estimador del tipo $\alpha X + (1 - \alpha)E[X|Y]$. Sin embargo, por la ecuación (8.3), el mejor estimador de este tipo tiene $\alpha = \alpha^*$, donde

$$\alpha^* = \frac{\text{Var}(E[X|Y]) - \text{Cov}(X, E[X|Y])}{\text{Var}(X) + \text{Var}(E[X|Y]) - 2 \text{Cov}(X, E[X|Y])}$$

Ahora mostraremos que $\alpha^* = 0$ y que la combinación de los estimadores X y $E[X|Y]$ no mejora lo que se obtiene si se emplea solamente $E[X|Y]$.

Observe primero que

$$\begin{aligned} \text{Var}(E[X|Y]) &= E[(E[X|Y])^2] - (E[E[X|Y]])^2 \\ &= E[(E[X|Y])^2] - (E[X])^2 \end{aligned} \quad (8.5)$$

Por otro lado,

$$\begin{aligned} \text{Cov}(X, E[X|Y]) &= E[XE[X|Y]] - E[X]E[E[X|Y]] \\ &= E[XE[X|Y]] - (E[X])^2 \\ &= E[E[XE[X|Y]|Y]] - (E[X])^2 \\ &\quad \text{(condicionando sobre } Y) \\ &= E[E[X|Y]E[X|Y]] - (E[X])^2 \\ &\quad \text{(pues dado } Y, E[X|Y] \text{ es constante)} \\ &= \text{Var}(E[X|Y]) \quad [\text{de (8.5)}] \end{aligned}$$

Así, vemos que no se logra una reducción de varianza adicional combinando los estimadores X y $E[X|Y]$.

Ahora ilustramos el uso del "condicionamiento" mediante varios ejemplos.

Ejemplo 8k Reconsideremos nuestro uso de la simulación para estimar π . En el ejemplo 3a del capítulo 3, mostramos la forma de estimar π determinando la frecuencia con la cual un punto elegido al azar en el cuadrado de área 4 con centro en el

origen cae dentro del círculo inscrito de radio 1. Específicamente, si $V_i = 2U_i - 1$, donde U_i , $i = 1, 2$, son números aleatorios, y sea

$$I = \begin{cases} 1 & \text{si } V_1^2 + V_2^2 \leq 1 \\ 0 & \text{en caso contrario} \end{cases}$$

entonces, como observamos en el ejemplo 3a, $E[I] = \pi/4$.

El uso del promedio de valores sucesivos de I para estimar $\pi/4$ se puede mejorar utilizando $E[I|V_1]$ en vez de I . Ahora,

$$\begin{aligned} E[I|V_1 = v] &= P\{V_1^2 + V_2^2 \leq 1 | V_1 = v\} \\ &= P\{v^2 + V_2^2 \leq 1 | V_1 = v\} \\ &= P\{V_2^2 \leq 1 - v^2\} \quad \text{por la independencia de } V_1 \text{ y } V_2 \\ &= P\{-(1 - v^2)^{1/2} \leq V_2 \leq (1 - v^2)^{1/2}\} \\ &= \int_{-(1-v^2)^{1/2}}^{(1-v^2)^{1/2}} \left(\frac{1}{2}\right) dx \quad \text{pues } V_2 \text{ es uniforme en } (-1, 1) \\ &= (1 - v^2)^{1/2} \end{aligned}$$

Por lo tanto,

$$E[I|V_1] = (1 - V_1^2)^{1/2}$$

de modo que el estimador $(1 - V_1^2)^{1/2}$ también tiene media $\pi/4$ y tiene menor varianza que I . Como

$$\begin{aligned} E[(1 - V_1^2)^{1/2}] &= \int_{-1}^1 (1 - x^2)^{1/2} \left(\frac{1}{2}\right) dx \\ &= \int_0^1 (1 - x^2)^{1/2} dx \\ &= E[(1 - U^2)^{1/2}] \end{aligned}$$

podemos simplificar un poco utilizando el estimador $(1 - U^2)^{1/2}$, donde U es un número aleatorio.

Para determinar con facilidad la mejora en la varianza obtenida nos valemos del estimador $(1 - U^2)^{1/2}$ en relación con el estimador I .

$$\begin{aligned} \text{Var}[(1 - U^2)^{1/2}] &= E[1 - U^2] - \left(\frac{\pi}{4}\right)^2 \\ &= \frac{2}{3} - \left(\frac{\pi}{4}\right)^2 = 0.0498 \end{aligned}$$

donde la primera desigualdad utiliza la identidad $\text{Var}(W) = E[W^2] - (E[W])^2$. Por otro lado, como I es una variable aleatoria Bernoulli con media $\pi/4$, tenemos

$$\text{Var}(I) = \left(\frac{\pi}{4}\right)\left(1 - \frac{\pi}{4}\right) = 0.1686$$

lo cual muestra que el condicionamiento produce una reducción de 70.44 por ciento en la varianza (además, sólo se necesita uno y no dos números aleatorios para cada ejecución de simulación, aunque esto se paga con cierto costo computacional por el cálculo de una raíz cuadrada).

Como es claro que la función $(1 - u^2)^{1/2}$ es una función monótona decreciente de u en la región $0 < u < 1$, esto implica que el estimador $(1 - U^2)^{1/2}$ mejora utilizando variables antitéticas. Es decir, el estimador

$$\frac{1}{2}[(1 - U^2)^{1/2} + (1 - (1 - U)^2)^{1/2}]$$

tiene menor varianza que $\frac{1}{2}[(1 - U_1^2)^{1/2} + (1 - U_2^2)^{1/2}]$.

Otra forma de mejorar al estimador $(1 - U^2)^{1/2}$ es mediante una variable de control. Una variable de control natural en este caso es U^2 y, como $E[U^2] = \frac{1}{3}$, podríamos emplear un estimador del tipo

$$(1 - U^2)^{1/2} + c\left(U^2 - \frac{1}{3}\right)$$

Podemos estimar la mejor c (a saber, $c^* = -\text{Cov}[(1 - U^2)^{1/2}, U^2]/\text{Var}(U^2)$) realizando la simulación para estimar el término de covarianza (también podríamos tomar U como variable de control; esto es diferente, pues una correlación entre dos variables aleatorias es tan sólo una medida de su "dependencia lineal" y no de su total dependencia, pero el uso de U^2 da una mejora mayor; véase el ejercicio 16). ■

Ejemplo 8l Revisión de la función de fiabilidad. Suponga, como en el ejemplo 8b, que $S_j, j = 1, \dots, n$ son independientes, con

$$P\{S_j = 1\} = 1 - P\{S_j = 0\} = p_j$$

y suponga que estamos interesados en estimar la función de fiabilidad $E[\phi(S_1, \dots, S_n)]$, donde ϕ es una función binaria creciente (es decir, sus únicos valores posibles son 0 y 1). Si generamos los valores de S_1, \dots, S_n , entonces una mejora sobre el

estimador en bruto $\phi(S_1, \dots, S_n)$ consiste en considerar la esperanza condicional dadas todas las S_j excepto una. Es decir, para i fijo,

$$E[\phi(S_1, \dots, S_n) | S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_n] \quad (8.6)$$

será un mejor estimador. (Para utilizar el estimador anterior simularíamos, en cualquier ejecución de simulación, todos los estados componentes S_j excepto S_i . La estimación en esa ejecución tomaría uno de tres valores posibles; será igual a 1 si dados los estados $S_j, j \neq i$, el sistema continúa funcionando aunque no funcione el componente i , o será igual a 0 si el sistema no funciona aunque lo hiciera el componente i , o será igual a p_i si el sistema funciona si i funciona y falla en caso contrario.)

El estimador anterior, dado por (8.6), se puede mejorar aún más gracias a variables antitéticas (pues el estimador sigue siendo una función monótona) o controlando mediante $\sum_{j \neq i} S_j$ el número de componentes simulados que funcionan. ■

Ejemplo 8m Un modelo de línea de espera con capacidad finita. Considere un sistema de línea de espera en el que las llegadas entran al sistema sólo si hay menos de otros N clientes en el sistema cuando alguien llega. Cualquier cliente que encuentra otros N al llegar se considera perdido para el sistema. Suponga además que los clientes potenciales llegan de acuerdo con un proceso Poisson con razón λ y que estamos interesados en el uso de la simulación para estimar la cantidad esperada de clientes perdidos hasta cierto instante fijo t .

Una ejecución de simulación constaría de la simulación del sistema anterior hasta el tiempo t . Si, para una ejecución dada, L denota el número de clientes perdidos, entonces el valor promedio de L , sobre todas las ejecuciones de simulación, es el estimador de simulación (en bruto) de la cantidad deseada $E[L]$. Sin embargo, podemos mejorar este estimador condicionando la cantidad de tiempo que el sistema está a toda su capacidad. Es decir, en vez de utilizar L , el número real de clientes perdidos hasta el instante t , consideramos $E[L | T_C]$, donde T_C es la cantidad total de tiempo en el intervalo $(0, t)$ en que hay N clientes en el sistema. Como los clientes siempre llegan con la razón de Poisson λ sin importar lo que ocurra en el sistema, tenemos que

$$E[L | T_C] = \lambda T_C$$

Por lo tanto se obtiene un estimador mejorado averiguando, para cada ejecución, el tiempo total en que hay N clientes en el sistema, digamos que $T_{C,i}$ es el tiempo que el sistema está a toda su capacidad durante la i -ésima ejecución. Entonces, el estimador mejorado de $E[L]$ es $\lambda \sum_{i=1}^k T_{C,i}/k$, donde k es el número de ejecuciones de simulación. De hecho, como el número esperado de clientes perdidos, dado el tiempo T_C a máxima capacidad, es precisamente λT_C , lo que hace este estimador es utilizar la esperanza

condicional real en vez de simular (e incrementar la varianza del estimador) una variable aleatoria Poisson con esta media.

Si el proceso de llegada fuese un proceso Poisson no homogéneo con función de intensidad $\lambda(s)$, $0 \leq s \leq t$, entonces no podríamos calcular la cantidad esperada condicional de clientes perdidos si sólo tuviéramos el tiempo total a capacidad máxima. Lo que ahora necesitamos son los tiempos reales en los que el sistema estaba a toda su capacidad. Así, condicionaremos los intervalos durante los cuales el sistema estaba a toda su capacidad. Si N_C denota el número de intervalos de $(0, t)$ durante los cuales el sistema está a toda su capacidad y si estos intervalos se designan como I_1, \dots, I_{N_C} , entonces

$$E[L|N_C, I_1, \dots, I_{N_C}] = \sum_{i=1}^{N_C} \int_{I_i} \lambda(s) ds$$

El uso del valor promedio de la cantidad anterior, sobre todas las ejecuciones de simulación, conduce a un mejor estimador de $E[L]$ que el estimador de simulación en bruto del número promedio perdido por ejecución, en el sentido de que tiene un error cuadrático medio menor.

Es posible combinar esto con otras técnicas de reducción de varianza al estimar $E[L]$. Por ejemplo, si M denota el número de clientes que realmente han entrado a servicio hasta el instante t , y $N(t)$ igual al número de llegadas hasta el instante t , entonces tenemos

$$N(t) = M + L$$

Al obtener las esperanzas de ambos lados tenemos que

$$\int_0^t \lambda(s) ds = E[M] + E[L]$$

Por lo tanto, $\int_0^t \lambda(s) ds - M$ es también un estimador insesgado de $E[L]$, lo que sugiere el uso del estimador combinado

$$\alpha \sum_{i=1}^{N_C} \int_{I_i} \lambda(s) ds + (1 - \alpha) \left(\int_0^t \lambda(s) ds - M \right)$$

Aquí se utiliza el valor de α dado por la ecuación (8.3), el cual puede estimarse mediante la simulación. ■

Ejemplo 8n Suponga que queremos estimar la suma esperada de los tiempos en el sistema para los primeros n clientes en un sistema de línea de espera. Es decir, si W_i es el tiempo que el i -ésimo cliente pasa en el sistema, nos interesa estimar

$$\theta = E \left[\sum_{i=1}^n W_i \right]$$

Sea S_i el "estado del sistema" en el momento en que llega el i -ésimo cliente, y consideremos el estimador

$$\sum_{i=1}^n E[W_i|S_i]$$

Como

$$E \left[\sum_{i=1}^n E[W_i|S_i] \right] = \sum_{i=1}^n E[E[W_i|S_i]] = \sum_{i=1}^n E[W_i] = \theta$$

tenemos que se trata de un estimador insesgado de θ . Se puede mostrar[†] que, en una amplia clase de modelos, este estimador tiene una varianza menor que la del estimador por simulación en bruto $\sum_{i=1}^n W_i$ (debemos observar que aunque es inmediato que $E[W_i|S_i]$ tiene menor varianza que W_i , esto no implica que $\sum_{i=1}^n E[W_i|S_i]$ tenga menor varianza que $\sum_{i=1}^n W_i$ debido a los términos de covarianza).

La cantidad S_i , que se refiere al estado del sistema visto por el i -ésimo cliente a su llegada, se supone que representa la menor cantidad de información que nos permite calcular el tiempo esperado condicional que el cliente invierte en el sistema. Por ejemplo, si hay un único servidor y todos los tiempos de servicio son exponenciales con media μ , entonces S_i hace referencia a N_i , el número de clientes en el sistema, hasta la i -ésima llegada. En este caso,

$$E[W_i|S_i] = E[W_i|N_i] = (N_i + 1)\mu$$

lo cual es consecuencia del hecho de que la i -ésima llegada tendrá que esperar N_i tiempos de servicio (uno de los cuales es la conclusión del servicio del cliente al que se atiende cuando el cliente i llega; pero, por la propiedad de la exponencial de no tener memoria, ese tiempo restante también será exponencial con media μ), todos los cuales tienen media μ , y luego debemos sumar a esto su propio tiempo de servicio. Así, el estimador que considera el valor promedio de la cantidad $\sum_{i=1}^n (N_i + 1)\mu$, sobre todas las ejecuciones de simulación es un mejor estimador que el valor promedio de $\sum_{i=1}^n W_i$. ■

ESTIMACIÓN DEL NÚMERO ESPERADO DE RENOVACIONES HASTA EL INSTANTE t

Suponga que ocurren "eventos" de manera aleatoria en el tiempo. Sea T_1 el tiempo del primer evento, T_2 el tiempo entre el primero y el segundo eventos, y, en general, T_n es

[†] S. M. Ross, "Simulating Average Delay-Variance Reduction by Conditioning", en *Probability in the Engineering and Informational Sciences*, vol. 2, núm. 3, 1988.

el tiempo entre el $(n - 1)$ -ésimo y el n -ésimo evento, $n \geq 1$. Si hacemos

$$S_n = \sum_{i=1}^n T_i$$

el primer evento ocurre en el instante S_1 , el segundo en el instante S_2 y, en general, el evento n -ésimo ocurre en el instante S_n (véase la figura 8.2). Sea $N(t)$ el número de eventos que ocurren hasta el instante t ; es decir, $N(t)$ es la máxima n para la cual el n -ésimo evento ocurre hasta el instante t , o bien en forma equivalente,

$$N(t) = \text{Máx}\{n: S_n \leq t\}$$

Si los tiempos entre los eventos T_1, T_2, \dots son independientes e idénticamente distribuidos, de acuerdo con cierta función de distribución F , entonces el proceso $\{N(t), t \geq 0\}$ se llama *proceso de renovación*.

Podemos simular fácilmente un proceso de renovación generando los tiempos entre las llegadas. Suponga ahora que queremos utilizar simulación para estimar $\theta = E[N(t)]$, el número medio de eventos hasta cierto tiempo fijo t . Para esto, simulamos de manera sucesiva los tiempos entre los eventos, manteniendo un registro con su suma (la cual representa los tiempos en los cuales ocurren los eventos) hasta que esta suma excede a t . Es decir, continuamos generando tiempos entre eventos hasta llegar al primer tiempo de evento posterior a t . Si $N(t)$ (el estimador de simulación en bruto) denota el número de eventos simulados hasta el instante t , vemos que una cantidad natural para utilizarse como variable de control es la sucesión de $N(t) + 1$ tiempos entre eventos ya generados. Es decir, si μ denota el tiempo medio entre eventos, entonces, como las variables aleatorias $T_i - \mu$ tienen media cero, se concluye que

$$E\left[\sum_{i=1}^{N(t)+1} (T_i - \mu)\right] = 0$$

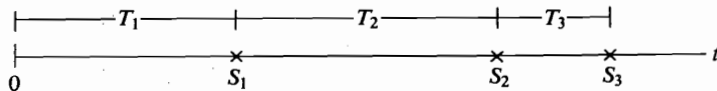


Figura 8.2 $x = \text{evento}$

Por lo tanto, podemos controlar mediante un estimador del tipo

$$\begin{aligned} N(t) + c \left[\sum_{i=1}^{N(t)+1} (T_i - \mu) \right] &= N(t) + c \left[\sum_{i=1}^{N(t)+1} T_i - \mu(N(t) + 1) \right] \\ &= N(t) + c[S_{N(t)+1} - \mu N(t) - \mu] \end{aligned}$$

Ahora, como S_n representa el tiempo del n -ésimo evento y $N(t) + 1$ representa el número de eventos hasta el instante t más 1, se tiene que $S_{N(t)+1}$ representa el tiempo del primer evento posterior al instante t . Por lo tanto, si $Y(t)$ denota el tiempo de t hasta el siguiente evento [$Y(t)$ se llama por lo general el exceso de vida en t], entonces

$$S_{N(t)+1} = t + Y(t)$$

de modo que el estimador controlado anterior se puede escribir

$$N(t) + c[t + Y(t) - \mu N(t) - \mu]$$

La mejor c está dada por

$$c^* = - \frac{\text{Cov}[N(t), Y(t) - \mu N(t)]}{\text{Var}[Y(t) - \mu N(t)]}$$

Para t grande, se puede mostrar que los términos relacionados con $N(t)$ dominan, pues su varianza crecerá linealmente con t , mientras que los demás términos permanecen acotados, de modo que para t grande

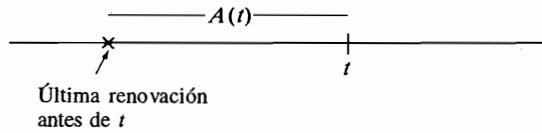
$$c^* \approx - \frac{\text{Cov}[N(t), -\mu N(t)]}{\text{Var}[-\mu N(t)]} = \frac{\mu \text{Var}[N(t)]}{\mu^2 \text{Var}[N(t)]} = \frac{1}{\mu}$$

Así, para t grande, el mejor estimador controlado del tipo anterior es cercano a

$$N(t) + \frac{1}{\mu} (t + Y(t) - \mu N(t) - \mu) = \frac{Y(t)}{\mu} + \frac{t}{\mu} - 1 \quad (8.7)$$

En otras palabras, para t grande, el valor crítico por determinar a partir de la simulación es $Y(t)$, el tiempo que va de t hasta la siguiente renovación.

El estimador anterior se puede mejorar utilizando el "condicionamiento". A saber, en vez de utilizar el tiempo observado real del primer evento posterior a t , podemos condicionar sobre $A(t)$, el tiempo en t desde el último evento (véase la figura 8.3). La cantidad $A(t)$ recibe el nombre de edad del proceso de renovación en t . [Si imaginamos un sistema que consta de un único elemento, el cual funciona para un tiempo aleatorio con distribución F , que luego falla y es reemplazado de inmediato por un nuevo elemento, entonces tenemos un proceso de renovación donde cada evento corresponde a la falla de un elemento. Entonces, la variable $A(t)$ se refiere a la edad del

Figura 8.3 Edad en t

elemento en uso en el instante t , donde por edad entendemos el tiempo que ha permanecido en uso.]

Ahora, si la edad del proceso en el instante t es x , la vida restante esperada del elemento es simplemente la cantidad esperada de exceso de un tiempo entre eventos con respecto de x , dado que es mayor que x . Es decir,

$$\begin{aligned} E[Y(t)|A(t) = x] &= E[T - x | T > x] \\ &= \int_x^\infty (y - x) \frac{f(y)}{1 - F(x)} dy \\ &\equiv \mu[x] \end{aligned}$$

donde lo anterior supone que F es una distribución continua con función de densidad f . Por lo tanto, si $\mu[x]$ se define como $E[T - x | T > x]$, vemos que

$$E[Y(t)|A(t)] = \mu[A(t)]$$

Así, para t grande, un mejor estimador de $E[N(t)]$ que el dado en la ecuación (8.7) es

$$\frac{\mu[A(t)]}{\mu} + \frac{t}{\mu} - 1 \quad (8.8)$$

8.4 Muestreo estratificado

Considere un sistema de línea de espera con capacidad finita, con la propiedad de que se vacía al final de cada día y comienza desde cero al principio del día siguiente. Los clientes llegan diariamente de acuerdo con un proceso Poisson, pero suponga que la razón de ese proceso puede cambiar cada día. Específicamente, suponga que para cualquier día dado, independientemente de lo ocurrido en días anteriores, la tasa de

llegada Poisson, es λ_1, λ_2 o λ_3 , cada una con probabilidad $\frac{1}{3}$. Nos interesa simular este sistema para estimar el número promedio de clientes perdidos (pues han llegado cuando el sistema estaba a toda su capacidad) durante un día.

Para simular lo anterior, comenzamos cada ejecución generando una variable aleatoria I que tome los valores 1, 2 o 3 con la misma probabilidad. Luego simulamos un día del sistema anterior bajo la hipótesis de que los clientes llegan de acuerdo con un proceso Poisson con razón λ_i . Sin embargo, desde el punto de vista intuitivo, como sabemos que cada día tiene la misma probabilidad de tener llegadas Poisson con la razón λ_i , $i = 1, 2, 3$, parecería preferible realizar exactamente una tercera parte de nuestras ejecuciones de simulación utilizando cada una de estas tres razones.

Para verificar esta idea intuitiva, sea L_i^* , $i = 1, 2, 3$, el número de clientes perdidos en un día, dado que las llegadas durante ese día son Poisson con media λ_i . Además, sea L el número de clientes perdidos en un día. Entonces, al condicionar sobre la tasa diaria de llegadas, vemos que

$$E[L] = \frac{1}{3} (E[L_1^*] + E[L_2^*] + E[L_3^*]) \quad (8.9)$$

Sea L_i el número de clientes perdidos el día i , con la misma distribución que L . Ahora justificaremos que

$$\text{Var}(L_1^* + L_2^* + L_3^*) \leq \text{Var}(L_1 + L_2 + L_3) \quad (8.10)$$

Como por (8.9) $(L_1^* + L_2^* + L_3^*)/3$ es un estimador insesgado de $E[L]$, (8.10) implica que el uso de cada λ_i en exactamente una tercera parte de las ejecuciones de simulación (opuesto a elegir aleatoriamente al principio de cada ejecución) produce un estimador insesgado con menor varianza.

Para demostrar la desigualdad (8.10), nos servimos de la fórmula de la varianza condicional

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

Como ambos términos de la derecha son no negativos, vemos que

$$\text{Var}(X) \geq E[\text{Var}(X|Y)]$$

Por lo tanto, si L representa el número de clientes perdidos en un día e I es el índice de la tasa de llegadas Poisson para ese día, vemos que

$$\begin{aligned} \text{Var}(L) &\geq E[\text{Var}(L|I)] \\ &= \sum_{i=1}^3 \text{Var}(L|I = i) P\{I = i\} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{3} \sum_{i=1}^3 \text{Var}(L_i^*) \\
 &= \frac{1}{3} \text{Var}(L_1^* + L_2^* + L_3^*)
 \end{aligned}$$

lo cual establece (8.10), pues $\text{Var}(L_1 + L_2 + L_3) = 3 \text{Var}(L)$.

Ejemplo 8o En días buenos, los clientes llegan a la cola infinita de un servidor de acuerdo con un proceso Poisson con tasa de 12 por hora, mientras que en otros días llegan de acuerdo con un proceso Poisson con tasa de cuatro por hora. Los tiempos de servicio, de todos los días, se distribuyen exponencialmente con tasa de uno por hora. Cada día, a las 10 horas, el sistema se apaga y todos los presentes en el servicio deben salir del sistema sin concluir su atención. Suponga que cada día es, de manera independiente, un día bueno con probabilidad 0.5 y que queremos utilizar la simulación para estimar θ , el número medio de clientes por día cuyo servicio no concluye.

Sea L el número de clientes cuyo servicio no se concluye en un día elegido al azar y sean L_g y L_o el número en un día bueno y en un día común, respectivamente. Se puede mostrar que L_g y L_o son variables aleatorias Poisson con medias respectivas $12(1 - e^{-10})$ y $4(1 - e^{-10})$. Por lo tanto, como la varianza de una variable aleatoria Poisson es igual a su media, vemos que

$$\text{Var}(L_g) \approx 12, \quad \text{Var}(L_o) \approx 4$$

Además, como L es igual a L_g con probabilidad $\frac{1}{2}$ y es igual a L_o con probabilidad $\frac{1}{2}$, se tiene que

$$\theta = E[L] = \frac{1}{2} E[L_g] + \frac{1}{2} E[L_o] \approx 8$$

$$E[L^2] = \frac{1}{2} E[L_g^2] + \frac{1}{2} E[L_o^2] \approx \frac{(12 + 144 + 4 + 16)}{2} = 88$$

donde empleamos la identidad $E[X^2] = \text{Var}(X) + (E[X])^2$. Por lo tanto,

$$\text{Var}(L) \approx 88 - 64 = 24$$

Así, si primero simulamos L_g y luego L_o , entonces la varianza del estimador $(L_g + L_o)/2$ es

$$\text{Var}\left(\frac{L_g + L_o}{2}\right) = \frac{[\text{Var}(L_g) + \text{Var}(L_o)]}{4} \approx 4$$

Por otro lado, si simulamos dos iteraciones de L , entonces la varianza del estimador $(L_1 + L_2)/2$ es

$$\text{Var}[(L_1 + L_2)/2] \approx \frac{24}{2} = 12$$

que es mayor que la varianza del estimador anterior por un factor aproximadamente igual a 3. ■

En el caso más general, suponga que las posibles tasas de llegada cada día son λ_i , con probabilidades respectivas p_i , $\sum_{i=1}^k p_i = 1$. En este caso, tendríamos que

$$E[L] = \sum_{i=1}^k E[L_i^*] p_i$$

donde L_i^* es el número perdido en un día donde los clientes llegan con la razón λ_i . Por lo tanto, en vez de elegir al azar una tasa de llegada cada día, un argumento similar al ya presentado muestra que es mejor realizar $100p_i\%$ de las ejecuciones con la tasa de llegada λ_i , $i = 1, \dots, k$. Si \bar{L}_i^* es el número promedio de clientes perdidos en las ejecuciones que utilizan la tasa de llegada λ_i , entonces el estimador de $\theta = E[L]$ con base en n ejecuciones es

$$\hat{\theta} = \sum_{i=1}^k p_i \bar{L}_i^*$$

La varianza del estimador está dada por

$$\text{Var}(\hat{\theta}) = \sum_{i=1}^k \frac{p_i^2 \text{Var}(L_i^*)}{n_i} \quad (8.11)$$

donde $n_i = np_i$ es el número de ejecuciones que usan la tasa de llegada λ_i . Si S_i^2 denota la varianza muestral del número de clientes perdidos en los días que utilizan la tasa de llegada λ_i , entonces $\text{Var}(\hat{\theta})$ se puede estimar como $\sum_{i=1}^k p_i^2 S_i^2 / n_i$.

Aunque la realización de np_i de las n ejecuciones con la tasa de llegada λ_i , $i = 1, \dots, k$, es mejor que elegir al azar la tasa al principio de cada día, no necesariamente es el número óptimo de ejecuciones por realizar empleando cada una de estas tasas. Suponga que planeamos realizar n ejecuciones de simulación. Si n_i es el número de estas ejecuciones que utilizan la tasa de llegada λ_i , $i = 1, \dots, k$, entonces el estimador de $E[L]$ será $\sum_{i=1}^k p_i \bar{L}_i^*$ con la varianza dada por (8.11). Aunque al principio no se conozcan las varianzas de L_i^* , $i = 1, \dots, k$, podríamos realizar un pequeño estudio de

simulación para estimarlas; digamos que utilizamos los estimadores s_i^2 . Entonces podríamos elegir n_i resolviendo el siguiente problema de optimización:

$$\begin{aligned} &\text{elegir } n_1, \dots, n_k, \quad \sum_{i=1}^k n_i = n \quad \text{para} \\ &\text{minimizar } \sum_{i=1}^k p_i^2 s_i^2 / n_i \end{aligned}$$

Una vez elegidas las n_i , estimaríamos $E[L]$ mediante $\sum_{i=1}^k p_i \bar{L}_i^*$ y estimaríamos su varianza por medio de $\sum_{i=1}^k p_i^2 S_i^2 / n_i$ donde, como antes, S_i^2 es la varianza muestral para las n_i ejecuciones que utilizan λ_i .

Por ejemplo, suponga que $k = 2$ y que realizamos nx ejecuciones con λ_1 y $n(1-x)$ ejecuciones con λ_2 . Entonces, si $p = p_1$, quisiéramos elegir x , $0 \leq x \leq 1$, que minimice

$$\frac{ps_1^2}{nx} + \frac{(1-p)s_2^2}{n(1-x)}$$

lo cual se logra mediante el cálculo diferencial. Si $p = \frac{1}{2}$, $s_1^2 = 12$, $s_2^2 = 4$, como en el ejemplo 8n, al resolver lo anterior obtenemos $x \approx 0.634$ el valor que minimiza la suma, lo cual muestra que debemos hacer aproximadamente 63.4 por ciento de las ejecuciones con $\lambda = 12$ y 36.6 por ciento con $\lambda = 4$.

En general, si queremos estimar $E[X]$ en una situación en la que X depende de una variable aleatoria S que toma alguno de los valores $1, \dots, k$ con probabilidades conocidas, entonces la técnica de estratificación de las ejecuciones de simulación en k grupos, donde el i -ésimo grupo satisface $S = i$, donde \bar{X}_i es el valor promedio de X en las ejecuciones tales que $S = i$, para luego estimar $E[X]$ $\sum_{i=1}^k E[X|S=i] P\{S=i\}$ mediante $\sum_{i=1}^k \bar{X}_i P\{S=i\}$, se llama *muestreo estratificado*.

Observación Es interesante notar que la demostración de que el muestreo estratificado conduce a una reducción de la varianza se vale de la fórmula de varianza condicional para afirmar que

$$\text{Var}(X) \geq E[\text{Var}(X|Y)]$$

mientras que la demostración de que el "condicionamiento" siempre reduce la varianza de un estimador (véase la sección 8.3) utiliza la fórmula de varianza condicional para afirmar que

$$\text{Var}(X) \geq \text{Var}(E[X|Y])$$

Como otra ilustración del muestreo estratificado, suponga que queremos realizar n ejecuciones de simulación para estimar

$$\theta = E[h(U)] = \int_0^1 h(x) dx$$

Si hacemos

$$S = j \quad \text{si} \quad \frac{j-1}{n} \leq U < \frac{j}{n}, \quad j = 1, \dots, n$$

entonces

$$\begin{aligned} \theta &= \frac{1}{n} \sum_{j=1}^n E[h(U)|S=j] \\ &= \frac{1}{n} \sum_{j=1}^n E[h(U_{(j)})] \end{aligned}$$

donde $U_{(j)}$ es uniforme en $((j-1)/n, j/n)$. Así, por lo anterior, más que generar U_1, \dots, U_n y luego emplear $\sum_{j=1}^n h(U_j)/n$ para estimar θ , obtenemos un mejor estimador utilizando

$$\hat{\theta} = \frac{1}{n} \sum_{j=1}^n h\left(\frac{U_j + j - 1}{n}\right)$$

Ejemplo 8p En el ejemplo 8k mostramos que

$$\frac{\pi}{4} = E[\sqrt{(1-U)^2}]$$

Por lo tanto, para estimar π generamos U_1, \dots, U_n y empleamos el estimador

$$\text{est} = \frac{4}{n} \sum_{j=1}^n \sqrt{1 - [(U_j + j - 1)/n]^2}$$

De hecho, podemos mejorar lo anterior utilizando variables antitéticas para obtener el estimador

$$\hat{\pi} = \frac{2}{n} \sum_{j=1}^n \left(\sqrt{1 - [(U_j + j - 1)/n]^2} + \sqrt{1 - [(j - U_j)/n]^2} \right)$$

Una simulación mediante el estimador $\hat{\pi}$ arrojó los siguientes resultados:

n	$\hat{\pi}$
5	3.161211
10	3.148751
100	3.141734
500	3.141615
1000	3.141601
5000	3.141593

Cuando $n = 5000$, el estimador $\hat{\pi}$ es correcto hasta seis cifras decimales. ■

Suponga de nuevo que nos interesa estimar $\theta = E[X]$, donde X depende de la variable aleatoria S , la cual toma uno de los valores $1, 2, \dots, k$ con probabilidades respectivas $p_i, i = 1, \dots, k$. Entonces

$$E[X] = p_1 E[X|S=1] + p_2 E[X|S=2] + \dots + p_k E[X|S=k]$$

Si se conocen todas las cantidades $E[X|S=i]$ (es decir, si se conoce $E[X|S]$), pero no se conocen las p_i , entonces para estimar θ generamos el valor de S y a continuación usamos el estimador de la esperanza condicional $E[X|S]$. Por otro lado, si lo que se conoce son las p_i y podemos generar a partir de la distribución condicional de X dado el valor de S , entonces es posible realizar la simulación para obtener estimadores $\hat{E}[X|S=i]$ de las cantidades $E[X|S=i]$ y luego utilizar el estimador del muestreo estratificado $\sum_{i=1}^k p_i \hat{E}[X|S=i]$ para estimar $E[X]$. Cuando se conocen algunas de las p_i y algunas de las $E[X|S=i]$, cabe emplear una combinación de estos métodos.

Ejemplo 8q En el juego de video póquer, un jugador inserta una moneda en una máquina, la que le da una mano aleatoria de cinco cartas. Luego, se permite que el jugador descarte algunas cartas, que son reemplazadas por otras de las 47 restantes. Enseguida, el jugador recibe cierta cantidad según sus cartas finales. El siguiente es un esquema característico de pago:

Mano	Pago
Flor imperial	800
Flor corrida	50
Cuatro de un tipo (póquer)	25
Full house	8
Flor	5
Corrida	4
Tres de un tipo (tercia)	3
Dos pares	2
Par alto (jacks o mayores)	1
Cualquier otro caso	0

En lo anterior, una mano se caracteriza dentro de una categoría si es de ese tipo y no es de uno mayor. Es decir, por ejemplo, una flor significa cinco cartas del mismo palo que no sean consecutivas.

Considere una estrategia en la cual no se piden cartas adicionales (es decir, el jugador se "planta") si las cartas originales forman una corrida o algo mayor, y que siempre conserva todos los pares o tercias que recibe. Para una estrategia dada de este

tipo, sea X la ganancia del jugador en una sola mano y suponga que estamos interesados en estimar $\theta = E[X]$. En vez de simplemente utilizar X como el estimador, comenzaremos condicionando sobre el tipo de mano que recibe en primera instancia el jugador. R representa una flor imperial; S , una flor corrida; 4 , cuatro de un tipo; 3 , tres de un tipo; 2 , dos pares; 1 , un par alto; 0 , un par bajo; y "otro", el resto de las manos no mencionadas. Entonces tenemos

$$E[X] = E[X|R]P\{R\} + E[X|S]P\{S\} + E[X|4]P\{4\} + E[X|\text{full}]P\{\text{full}\} + E[X|\text{flor}]P\{\text{flor}\} + E[X|\text{corrida}]P\{\text{corrida}\} + E[X|3]P\{3\} + E[X|2]P\{2\} + E[X|1]P\{1\} + E[X|0]P\{0\} + E[X|\text{otro}]P\{\text{otro}\}$$

Ahora, si $C = \left(\frac{52}{5}\right)^{-1}$, tenemos

$$P\{R\} = 4C = 1.539 \times 10^{-6}$$

$$P\{S\} = 4 \cdot 9 \cdot C = 1.3851 \times 10^{-4}$$

$$P\{4\} = 13 \cdot 48 \cdot C = 2.40096 \times 10^{-4}$$

$$P\{\text{full}\} = 13 \cdot 12 \binom{4}{3} \binom{4}{2} C = 1.440576 \times 10^{-3}$$

$$P\{\text{flor}\} = 4 \left(\binom{13}{5} - 10 \right) C = 1.965402 \times 10^{-3}$$

$$P\{\text{corrida}\} = 10(4^5 - 4)C = 3.924647 \times 10^{-3}$$

$$P\{3\} = 13 \binom{12}{2} 4^3 C = 2.1128451 \times 10^{-2}$$

$$P\{2\} = \binom{13}{2} 44 \binom{4}{2} \binom{4}{2} C = 4.7539016 \times 10^{-2}$$

$$P\{1\} = 4 \binom{4}{2} \binom{12}{3} 4^3 C = 0.130021239$$

$$P\{0\} = 9 \binom{4}{2} \binom{12}{3} 4^3 C = 0.292547788$$

$$P\{\text{otro}\} = 1 - P\{R\} - P\{S\} - P\{\text{full}\} - P\{\text{flor}\} - P\{\text{corrida}\} - \sum_{i=0}^4 P\{i\} = 0.5010527$$

Por lo tanto, vemos que

$$E[X] = 0.0512903 + \sum_{i=0}^3 E[X|i]P\{i\} + E[X|\text{otro}]0.5010527$$

Podemos calcular $E[X|3]$ en forma analítica observando que las dos cartas nuevas vendrán de una baraja menor, de 47 cartas, la cual contiene una carta de una denomi-

nación (a saber, la denominación a la cual pertenecen las tres cartas de un tipo, ya obtenidas), tres cartas de dos denominaciones y cuatro de las otras 10 denominaciones. Así, si F es la mano final, tenemos que

$$P\{F = 4 \mid \text{dadas tres cartas}\} = \frac{46}{\binom{47}{2}} = 0.042553191$$

$$P\{F = \text{full} \mid \text{dadas tres cartas}\} = \frac{2 \cdot 3 + 10 \cdot 6}{\binom{47}{2}} = 0.061054579$$

$$P\{F = 3 \mid \text{dadas tres cartas}\} = 1 - 0.042553191 - 0.061054579 = 0.89639223$$

Por lo tanto,

$$\begin{aligned} E[X \mid 3] &= 25(0.042553191) + 8(0.061054579) + 3(0.89639223) \\ &= 4.241443097 \end{aligned}$$

De manera similar, podemos deducir en forma analítica (la deducción se deja como ejercicio) $E[X \mid i]$ para $i = 0, 1, 2$.

Al ejecutar la simulación, debemos generar una mano. Si ésta contiene al menos un par o una mano mayor, entonces debe descartarse e iniciar de nuevo el proceso. Cuando recibimos una mano que no contiene un par (o cualquier mano mayor), debemos utilizar la estrategia establecida por nosotros para descartar y recibir nuevas cartas. Si X_o es el pago por esta mano, entonces X_o es el estimador de $E[X \mid \text{otro}]$, y el estimador de $\theta = E[X]$ con base en esta única ejecución es

$$\begin{aligned} \hat{\theta} &= 0.0512903 + 0.021128451(4.241443097) + 0.047539016E[X \mid 2] \\ &\quad + 0.130021239E[X \mid 1] + 0.292547788E[X \mid 0] + 0.5010527X_o \end{aligned}$$

Observe que la varianza del estimador es

$$\text{Var}(\hat{\theta}) = (0.5010527)^2 \text{Var}(X_o)$$

Observaciones

1. Hemos supuesto que la estrategia utilizada siempre concuerda con una mano firme y siempre conserva todos los pares que se tengan. Sin embargo, para los pagos dados en este caso, no es una estrategia óptima. Por ejemplo, si uno recibe las cartas 2, 10, J, Q, K, todas de espadas, entonces en vez de quedarse con esta flor es mejor descartar el 2 y recibir otra carta (¿por qué?). Además, si se reciben las cartas 10, J, Q, K, todas de espadas, junto con el 10 de corazones, es mejor descartar el 10 de corazones y recibir una carta que conserve el par de 10.

2. Podríamos haber empleado el muestreo estratificado aún más, descomponiendo la categoría "otro" en, digamos, las "otras" manos que contienen cuatro cartas del mismo palo y aquellas que no. No es difícil calcular en forma analítica la probabilidad de que una mano no tenga un par pero sí tenga cuatro cartas del mismo palo. Entonces, utilizaríamos la simulación para estimar los pagos esperados condicionales en estos "otros" dos casos. ■

8.5 Muestreo de importancia

Sea $\mathbf{X} = (X_1, \dots, X_n)$ un vector de variables aleatorias con una función de densidad conjunta (o función de masa conjunta en el caso discreto) $f(\mathbf{x}) = f(x_1, \dots, x_n)$, y suponga que estamos interesados en estimar

$$\theta = E[h(\mathbf{X})] = \int h(\mathbf{x})f(\mathbf{x}) d\mathbf{x}$$

donde lo anterior es una integral n -dimensional (si las X_i son discretas, la integral se interpreta como una suma con n índices).

Suponga que una simulación directa del vector aleatorio \mathbf{X} , así como el cálculo de valores de $h(\mathbf{X})$, es ineficaz, pues (a) es difícil simular un vector aleatorio con función de densidad $f(\mathbf{x})$, o (b) la varianza de $h(\mathbf{X})$ es grande, o (c) una combinación de (a) y (b).

Otra forma en la cual se puede utilizar la simulación para estimar θ es observar que si $g(\mathbf{x})$ es otra densidad de probabilidad tal que $f(\mathbf{x}) = 0$ siempre que $g(\mathbf{x}) = 0$, entonces podemos expresar θ como

$$\begin{aligned} \theta &= \int \frac{h(\mathbf{x})f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \\ &= E_g \left[\frac{h(\mathbf{X})f(\mathbf{X})}{g(\mathbf{X})} \right] \end{aligned} \quad (8.12)$$

donde hemos escrito E_g para destacar que el vector aleatorio \mathbf{X} tiene densidad conjunta $g(\mathbf{x})$.

La ecuación (8.12) implica que θ se puede estimar mediante la generación sucesiva de valores de un vector aleatorio \mathbf{X} con función de densidad $g(\mathbf{x})$ y luego utilizar como estimador el promedio de los valores de $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$. Si se puede elegir una función de densidad $g(\mathbf{x})$ de modo que la variable aleatoria $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ tenga una varianza pequeña, entonces este método, conocido como *muestreo de importancia*, puede producir un estimador eficiente de θ .

Ahora intentaremos hacernos una idea del porqué el muestreo de importancia resulta útil. Para comenzar, observemos que $f(\mathbf{X})$ y $g(\mathbf{X})$ representan las posibilidades respectivas de obtener el vector \mathbf{X} cuando \mathbf{X} es un vector aleatorio con densidades respectivas f y g . Por lo tanto, si \mathbf{X} se distribuye de acuerdo con g , entonces por lo general ocurrirá que $f(\mathbf{X})$ será pequeño con respecto de $g(\mathbf{X})$, y así, cuando \mathbf{X} se simule de acuerdo con g , el cociente de verosimilitud $f(\mathbf{X})/g(\mathbf{X})$ será pequeño con respecto de 1. Sin embargo, es fácil verificar que su media es 1:

$$E_g \left[\frac{f(\mathbf{X})}{g(\mathbf{X})} \right] = \int \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} = \int f(\mathbf{x}) d\mathbf{x} = 1$$

Así, vemos que aunque $f(\mathbf{X})/g(\mathbf{X})$ por lo general es menor que 1, su media es igual a 1, lo cual implica que en ocasiones es grande, por lo cual tenderá a tener una varianza grande. Así, ¿cómo puede $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ tener una varianza pequeña? La respuesta es que a veces podemos elegir una densidad g tal que aquellos valores de \mathbf{x} para los cuales $f(\mathbf{x})/g(\mathbf{x})$ es grande sean precisamente los valores para los cuales $h(\mathbf{x})$ es demasiado pequeña, de modo que el cociente $h(\mathbf{X})f(\mathbf{X})/g(\mathbf{X})$ siempre sea pequeño. Como esto requiere que en ocasiones $h(\mathbf{x})$ sea pequeño, el muestreo de importancia parece funcionar mejor al estimar una probabilidad pequeña, ya que en este caso la función $h(\mathbf{x})$ es igual a 1 cuando \mathbf{x} está en algún conjunto y es igual a 0 en caso contrario.

Ahora consideraremos la cuestión de cómo elegir una densidad g adecuada. Veremos que las llamadas densidades transformadas son útiles. Sea $M(t) = E[e^{tX}] = \int e^{tx} f(x) dx$ la función generadora de momentos correspondiente a una densidad unidimensional f .

Definición Una función de densidad

$$f_t(x) = \frac{e^{tx} f(x)}{M(t)}$$

es una densidad transformada de f , $-\infty < t < \infty$.

Una variable aleatoria con densidad f_t tiende a ser mayor que una con densidad f cuando $t > 0$ y tiende a ser menor cuando $t < 0$.

En ciertos casos, las densidades transformadas f_t tienen la misma forma paramétrica que f .

Ejemplos 8r Si f es la densidad exponencial con razón λ , entonces

$$f_t(x) = C e^{tx} \lambda e^{-\lambda x} = C e^{-(\lambda-t)x}$$

donde $C = 1/M(t)$ no depende de x . Por lo tanto, para $t < \lambda$, f_t es una densidad exponencial con razón $\lambda - t$.

Si f es una función de masa de probabilidad Bernoulli con parámetro p , entonces

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Por lo tanto, $M(t) = E[e^{tX}] = pe^t + 1 - p$, y entonces

$$\begin{aligned} f_t(x) &= \frac{1}{M(t)} (pe^t)^x (1-p)^{1-x} \\ &= \left(\frac{pe^t}{pe^t + 1 - p} \right)^x \left(\frac{1-p}{pe^t + 1 - p} \right)^{1-x} \end{aligned}$$

Es decir, f_t es la función de masa de probabilidad de una variable aleatoria Bernoulli con parámetro $p_t = (pe^t)/(pe^t + 1 - p)$.

Dejamos como ejercicio mostrar que si f es una densidad normal con parámetros μ y σ^2 , entonces f_t es una densidad normal con media $\mu + \sigma^2 t$ y varianza σ^2 . ■

En ciertas situaciones, la cantidad de interés es la suma de las variables aleatorias independientes X_1, \dots, X_n . En este caso, la densidad conjunta f es el producto de densidades unidimensionales. Es decir,

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

donde f_i es la función de densidad de X_i . En esta situación, con frecuencia es útil generar las X_i de acuerdo con sus densidades transformadas, utilizando una elección común de t .

Ejemplo 8s Sean X_1, \dots, X_n variables aleatorias independientes con funciones de densidad de probabilidad (o masa) respectivas f_i , para $i = 1, \dots, n$. Suponga que estamos interesados en aproximar la probabilidad de que su suma sea al menos tan grande como a , donde a es mucho mayor que la media de la suma. Es decir, estamos interesados en

$$\theta = P\{S \geq a\}$$

donde $S = \sum_{i=1}^n X_i$, y donde $a > \sum_{i=1}^n E[X_i]$. Haciendo $I\{S \geq a\}$ igual a 1 si $S \geq a$ e igual a 0 en caso contrario, tenemos que

$$\theta = E_t[I\{S \geq a\}]$$

donde $\mathbf{f} = (f_1, \dots, f_n)$. Ahora suponga que simulamos X_i de acuerdo con la función

transformada de masa $f_{i,n}$, $i = 1, \dots, n$, con el valor de t , $t > 0$, aún por determinar. El estimador del muestreo de importancia de θ sería entonces

$$\hat{\theta} = I\{S \geq a\} \prod \frac{f_i(X_i)}{f_{i,t}(X_i)}$$

Ahora,

$$\frac{f_i(X_i)}{f_{i,t}(X_i)} = M_i(t)e^{-tX_i}$$

y entonces,

$$\hat{\theta} = I\{S \geq a\} M(t)e^{-tS}$$

donde $M(t) = \prod M_i(t)$ es la función generadora de momentos de S . Como $t > 0$ e $I\{S > a\}$ es igual a 0 cuando $S < a$, tenemos que

$$I\{S \geq a\}e^{-tS} \leq e^{-ta}$$

y entonces

$$\hat{\theta} \leq M(t)e^{-ta}$$

Para que la cota sobre el estimador sea lo más pequeña posible, elegimos t , $t > 0$, de modo que minimice $M(t)e^{-ta}$. Al hacer esto, obtendremos un estimador cuyo valor en cada iteración está entre 0 y $\min_t M(t)e^{-ta}$. Se puede mostrar que la t minimizante, que denotamos t^* , es tal que

$$E_{t^*}[S] = E_{t^*}\left[\sum_{i=1}^n X_i\right] = a$$

donde en lo anterior queremos indicar que el valor esperado se debe considerar bajo la hipótesis de que la distribución de X_i es f_{i,t^*} para $i = 1, \dots, n$.

Por ejemplo, suponga que X_1, \dots, X_n son variables aleatorias independientes de tipo Bernoulli con parámetros respectivos p_i , para $i = 1, \dots, n$. Luego, si generamos los X_i de acuerdo con sus funciones transformadas de masa $p_{i,t}$, $i = 1, \dots, n$, entonces el estimador del muestreo de importancia de $\theta = P\{S \geq a\}$ es

$$\hat{\theta} = I\{S \geq a\}e^{-tS} \prod_{i=1}^n (p_i e^t + 1 - p_i)$$

Como $p_{i,t}$ es la función de masa de una variable aleatoria Bernoulli con parámetro $(p_i e^t)/(p_i e^t + 1 - p_i)$, tenemos que

$$E_t\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \frac{p_i e^t}{p_i e^t + 1 - p_i}$$

El valor de t que hace lo anterior igual a a se puede aproximar en forma numérica y mediante la t utilizada en la simulación.

Como ilustración del método, suponga que $n = 20$, $p_i = 0.4$, $a = 16$. Entonces

$$E_t[S] = 20 \frac{0.4e^t}{0.4e^t + 0.6}$$

Al igualar esto a 16 y aplicar un poco de álgebra, se tiene que

$$e^{t^*} = 6$$

Así, si generamos las variables Bernoulli mediante el parámetro $(0.4e^{t^*})/(0.4e^{t^*} + 0.6) = 0.8$, entonces, como

$$M(t^*) = (0.4e^{t^*} + 0.6)^{20} \quad \text{y} \quad e^{-t^*S} = (1/6)^S$$

vemos que el estimador del muestreo de importancia es

$$\hat{\theta} = I\{S \geq 16\}(1/6)^{S/20}$$

Lo anterior implica que

$$\hat{\theta} \leq (1/6)^{16/20} = 81/2^{16} = 0.001236$$

Es decir, en cada iteración el valor del estimador está entre 0 y 0.001236. Como en este caso θ es la probabilidad de que una variable aleatoria binomial con parámetros 20, 0.4 sea al menos 16, esto se puede calcular de manera explícita, y el resultado da $\alpha = 0.000317$. Por lo tanto, el estimador de simulación en bruto I , que toma en cada iteración el valor 0 si la suma de las Bernoulli con parámetro 0.4 es menor que 16 y el valor 1 en caso contrario, tendrá varianza

$$\text{Var}(I) = \theta(1 - \theta) = 3.169 \times 10^{-4}$$

Por otro lado, el hecho de que $0 \leq \theta \leq 0.001236$ implica que (véase el ejercicio 22)

$$\text{Var}(\hat{\theta}) \leq 2.9131 \times 10^{-7}$$

Ejemplo 8t Considere una cola con un único servidor, en la cual los tiempos entre las llegadas sucesivas de los clientes tienen una función de densidad f y los tiempos de servicio tienen densidad g . Sea D_n la cantidad de tiempo que la n -ésima llegada espera en la fila y suponga que estamos interesados en estimar $\alpha = P\{D_n \geq a\}$ donde a es mucho mayor que $E[D_n]$. En vez de generar los sucesivos tiempos entre las llegadas y de servicio de acuerdo con f y g , respectivamente, debemos generarlos según las densidades f_{-t} y g_t , donde t es un número positivo por determinar. Observe que el uso de estas distribuciones y no de f y g produce menores tiempos entre las llegadas (pues $-t < 0$) y mayores tiempos de servicio. Por lo tanto, habrá una mayor probabilidad de que $D_n > a$

que si hubiésemos simulado mediante las densidades f y g . El estimador del muestreo de importancia de α sería entonces

$$\hat{\alpha} = I\{D_n > a\} e^{n(S_n - Y_n)} [M_f(-t) M_g(t)]^n$$

donde S_n es la suma de los primeros n tiempos entre llegadas, Y_n es la suma de los primeros n tiempos de servicio, y M_f y M_g son las funciones generadoras de momentos de las densidades f y g , respectivamente. El valor de t utilizado en este caso se debe determinar experimentando con diversas opciones. ■

Ejemplo 8u Sea X_1, X_2, \dots una sucesión de variables aleatorias normales, independientes e idénticamente distribuidas con media μ y varianza 1, donde $\mu < 0$. Un problema importante en la teoría de control de calidad (específicamente, en el análisis de los diagramas de suma acumulativa) consiste en determinar la probabilidad de que las sumas parciales de estos valores sean mayores que B antes de ser menores que $-A$. Es decir, sea

$$S_n = \sum_{i=1}^n X_i$$

y defina

$$N = \text{Mín} \{n : S_n < -A, \text{ o bien } S_n > B\}$$

donde A y B son números positivos fijos. Estamos interesados en estimar

$$\theta = P\{S_N > B\}$$

Una forma eficaz para estimar θ es la simulación de las X_i como si fueran normales con media $-\mu$ y varianza 1, y detenerse de nuevo cuando su suma es mayor que B o menor que $-A$ (como $-\mu$ es positiva, ocurre con más frecuencia que la suma sea mayor que B que cuando simulamos con la media negativa original). Si X_1, \dots, X_N denotan las variables simuladas (cada una normal con media $-\mu$ y varianza 1) y

$$I = \begin{cases} 1 & \text{si } \sum_{i=1}^N X_i > B \\ 0 & \text{en caso contrario} \end{cases}$$

entonces el estimador de θ a partir de esta ejecución es

$$I \prod_{i=1}^N \left[\frac{f_{\mu}(X_i)}{f_{-\mu}(X_i)} \right] \quad (8.13)$$

donde f_{μ} es la densidad normal con media μ y varianza 1. Como

$$\frac{f_{\mu}(x)}{f_{-\mu}(x)} = \frac{\exp\left\{-\frac{(x-\mu)^2}{2}\right\}}{\exp\left\{-\frac{(x+\mu)^2}{2}\right\}} = e^{2\mu x}$$

(8.13) implica que el estimador θ con base en esta ejecución es

$$I \exp\left\{2\mu \sum_{i=1}^N X_i\right\} = I \exp\{2\mu S_N\}$$

Cuando I es igual a 1, S_N excede a B y, como $\mu < 0$, en este caso el estimador es menor que $e^{2\mu B}$. Es decir, en vez de obtener en cada ejecución el valor 0 o 1 (como ocurriría con una simulación directa), obtenemos el valor 0 o un valor menor que $e^{2\mu B}$, lo cual revela el porqué este método de muestreo de importancia produce una menor varianza; por ejemplo, si $\mu = -0.1$ y $B = 5$, el estimador de cada ejecución está entre 0 y $e^{-1} = 0.3679$. Además, lo anterior es importante desde el punto de vista teórico, pues muestra que

$$P\{\text{cruzar } B \text{ antes que } -A\} \leq e^{2\mu B}$$

Como esto es verdadero para toda A positiva, obtenemos el interesante resultado

$$P\{\text{llegar a cruzar } B\} \leq e^{2\mu B} \quad \blacksquare$$

Ejemplo 8v Sea $\mathbf{X} = (X_1, \dots, X_{100})$ una permutación aleatoria de $(1, 2, \dots, 100)$. Es decir, \mathbf{X} tiene la misma probabilidad de ser cualquiera de las 100! permutaciones. Suponga que estamos interesados en utilizar simulación para estimar

$$\theta = P\left\{\sum_{j=1}^{100} jX_j > 290,000\right\}$$

Para tener una idea de la magnitud de θ , podemos comenzar calculando la media y la desviación estándar de $\sum_{j=1}^{100} jX_j$. De hecho, no es difícil mostrar que

$$E\left[\sum_{j=1}^{100} jX_j\right] = 100(101)^2/4 = 255,025$$

$$\text{DE}\left(\sum_{j=1}^{100} jX_j\right) = \sqrt{(99)(100)^2(101)^2/144} = 8374.478$$

Por lo tanto, si suponemos que $\sum_{j=1}^{100} jX_j$ se distribuye aproximadamente en forma normal y Z representa una variable aleatoria normal estándar, tenemos que

$$\begin{aligned}\theta &\approx P\left\{Z > \frac{290,000 - 255,025}{8374.478}\right\} \\ &= P\{Z > 4.1764\} \\ &= 0.00001481\end{aligned}$$

Así, es claro que θ es una probabilidad pequeña y vale la pena considerar un estimador del muestreo de importancia.

Para aplicar el muestreo de importancia, quisiéramos generar la permutación \mathbf{X} de modo que exista una probabilidad mucho mayor de que $\sum_{j=1}^{100} jX_j > 290,000$. De hecho, debemos intentar con una probabilidad cercana a 0.5. Ahora, $\sum_{j=1}^{100} jX_j$ alcanzará su máximo valor cuando $X_j = j$, $j = 1, \dots, 100$, y de hecho tenderá a ser grande cuando X_j tienda a ser grande cuando j sea grande y pequeño cuando j sea pequeño. Una forma de generar una permutación \mathbf{X} que tienda a ser de este tipo es la siguiente: generamos variables aleatorias exponenciales independientes Y_j , $j = 1, \dots, 100$, con razones respectivas λ_j , $j = 1, \dots, 100$, donde λ_j , $j = 1, \dots, 100$ es una sucesión creciente cuyos valores hay que especificar. Para $j = 1, \dots, 100$, sea X_j el índice del j -ésimo mayor de estos valores. Es decir,

$$Y_{X_1} > Y_{X_2} > \dots > Y_{X_{100}}$$

Como para j grande, Y_j tenderá a ser uno de los menores Y , entonces X_j tenderá a ser grande cuando j sea grande y así $\sum_{j=1}^{100} jX_j$ tenderá a ser mayor que si \mathbf{X} fuese una permutación distribuida de manera uniforme.

Calculemos $E[\sum_{j=1}^{100} jX_j]$. Para esto, sea $R(j)$ el rango de Y_j , $j = 1, \dots, 100$, donde rango 1 significa el más grande, rango 2 el segundo más grande, y así hasta el rango 100, que es el más pequeño. Observe que como X_j es el índice del j -ésimo más grande de los Y , $R(X_j) = j$. Por lo tanto,

$$\sum_{j=1}^{100} jX_j = \sum_{j=1}^{100} R(X_j)X_j = \sum_{j=1}^{100} jR(j)$$

donde la última igualdad se debe a que X_1, \dots, X_{100} es una permutación de $1, \dots, 100$. Por lo tanto, vemos que

$$E\left[\sum_{j=1}^{100} jX_j\right] = \sum_{j=1}^{100} jE[R(j)]$$

Para calcular $E[R_j]$, sea $I(i, j) = 1$ si $Y_j < Y_i$ e igual a 0 en otro caso; observe que

$$R_j = 1 + \sum_{i:i \neq j} I(i, j)$$

En palabras, la ecuación anterior establece que el rango de Y_j es 1 más el número de Y_i que son mayores que ella. Por lo tanto, al calcular las esperanzas y basarnos en el hecho de que

$$P\{Y_j < Y_i\} = \frac{\lambda_j}{\lambda_i + \lambda_j},$$

obtenemos que

$$E[R_j] = 1 + \sum_{i:i \neq j} \frac{\lambda_j}{\lambda_i + \lambda_j}$$

y así

$$E\left[\sum_{j=1}^{100} jX_j\right] = \sum_{j=1}^{100} j\left(1 + \sum_{i:i \neq j} \frac{\lambda_j}{\lambda_i + \lambda_j}\right)$$

Si hacemos $\lambda_j = j^{0.7}$, $j = 1, \dots, 100$, entonces un cálculo muestra que $E[\sum_{j=1}^{100} jX_j] = 290,293.6$, de modo que cuando \mathbf{X} sea generado mediante estos valores parecería que

$$P\left\{\sum_{j=1}^{100} jX_j > 290,000\right\} \approx 0.5$$

Así, sugerimos que para obtener el estimador de simulación primero deben generarse exponenciales independientes Y_j con razones respectivas $j^{0.7}$, y luego hacer que X_j sea el índice del j -ésimo más grande, $j = 1, \dots, 100$. Sea $I = 1$ si $\sum_{j=1}^{100} jX_j > 290,000$ e igual a 0 en caso contrario. Ahora, el resultado será \mathbf{X} cuando $Y_{X_{100}}$ sea el menor Y , $Y_{X_{99}}$ el segundo menor y así sucesivamente. La probabilidad de este resultado es $1/(100)!$ cuando \mathbf{X} puede tomar, con la misma probabilidad, cualquiera de las permutaciones, mientras que su probabilidad con la simulación desarrollada como antes es

$$\frac{(X_{100})^{0.7}}{\sum_{j=1}^{100} (X_j)^{0.7}} \frac{(X_{99})^{0.7}}{\sum_{j=1}^{99} (X_j)^{0.7}} \dots \frac{(X_2)^{0.7}}{\sum_{j=1}^2 (X_j)^{0.7}} \frac{(X_1)^{0.7}}{(X_1)^{0.7}}$$

Por lo tanto, el estimador del muestreo de importancia a partir de una única ejecución es

$$\hat{\theta} = \frac{I}{(100)!} \frac{\prod_{n=1}^{100} \left(\sum_{j=1}^n (X_j)^{0.7}\right)}{\left(\prod_{n=1}^{100} n\right)^{0.7}} = \frac{I \prod_{n=1}^{100} \left(\sum_{j=1}^n (X_j)^{0.7}\right)}{\left(\prod_{n=1}^{100} n\right)^{1.7}}$$

Antes de iniciar la simulación, hay que calcular los valores de $C = 1.7 \sum_{j=1}^{100} \log(n)$ y $a(j) = -j^{-0.7}$, $j = 1, \dots, 100$. Luego se puede obtener una ejecución de simulación como sigue:

Para $j = 1$ a 100

Generar un número aleatorio U

$Y_j = a(j) \log U$

Siguiente

Sea X_j , $j = 1, \dots, 100$, tal que Y_{X_j} es la j -ésima Y más grande

Si $\sum_{j=1}^n jX_j \leq 290,000$ hacer $\hat{\theta} = 0$ y terminar

$S = 0$, $P = 0$

Para $n = 1$ a 100

$S = S + (X_n)^{0.7}$

$P = P + \log(S)$

Siguiente

$\hat{\theta} = e^{P-C}$

Una muestra de 50,000 ejecuciones de simulación produjo la estimación $\hat{\theta} = 3.77 \times 10^{-6}$, con una varianza muestral 1.89×10^{-8} . Como la varianza del estimador de simulación en bruto, que es igual a 1 si $\sum_{j=1}^{100} jX_j > 290,000$ y es igual a 0 en caso contrario, es $\text{Var}(I) = \theta(1 - \theta) \approx 3.77 \times 10^{-6}$, vemos que

$$\frac{\text{Var}(I)}{\text{Var}(\hat{\theta})} \approx 199.47$$

El muestreo de importancia también es útil al estimar una esperanza condicional cuando se está condicionando sobre un evento raro. Es decir, suponga que \mathbf{X} es un vector aleatorio con función de densidad f y que estamos interesados en estimar

$$\theta = E[h(\mathbf{X}) | \mathbf{X} \in \mathcal{A}]$$

donde $h(\mathbf{x})$ es una función arbitraria con valores reales y $P\{\mathbf{X} \in \mathcal{A}\}$ es una probabilidad pequeña desconocida. Como la densidad condicional de \mathbf{X} dado que está en \mathcal{A} es

$$f(\mathbf{x} | \mathbf{X} \in \mathcal{A}) = \frac{f(\mathbf{x})}{P\{\mathbf{X} \in \mathcal{A}\}}, \quad \mathbf{x} \in \mathcal{A}$$

tenemos que

$$\begin{aligned} \theta &= \frac{\int_{\mathbf{x} \in \mathcal{A}} h(\mathbf{x}) f(\mathbf{x}) d(\mathbf{x})}{P\{\mathbf{X} \in \mathcal{A}\}} \\ &= \frac{E[h(\mathbf{X}) I(\mathbf{X} \in \mathcal{A})]}{E[I(\mathbf{X} \in \mathcal{A})]} \end{aligned}$$

$$= \frac{E[N]}{E[D]}$$

donde $E[N]$ y $E[D]$ se definen como iguales al numerador y al denominador de la expresión anterior; $I\{\mathbf{X} \in \mathcal{A}\}$ se define como 1 si $\mathbf{X} \in \mathcal{A}$ y 0 en caso contrario. Por lo tanto, en vez de simular \mathbf{X} de acuerdo con la densidad f , lo cual haría bastante improbable que estuviese en \mathcal{A} , la simulamos de acuerdo con alguna otra densidad g que haga este evento más probable. Si simulamos k vectores aleatorios $\mathbf{X}^1, \dots, \mathbf{X}^k$ de acuerdo con g , entonces podemos estimar $E[N]$ mediante $1/k \sum_{i=1}^k N_i$ y $E[D]$ mediante $1/k \sum_{i=1}^k D_i$ donde

$$N_i = \frac{h(\mathbf{X}^i) I(\mathbf{X}^i \in \mathcal{A}) f(\mathbf{X}^i)}{g(\mathbf{X}^i)}$$

y

$$D_i = \frac{I(\mathbf{X}^i \in \mathcal{A}) f(\mathbf{X}^i)}{g(\mathbf{X}^i)}$$

Así, obtenemos el siguiente estimador de θ :

$$\hat{\theta} = \frac{\sum_{i=1}^k h(\mathbf{X}^i) I(\mathbf{X}^i \in \mathcal{A}) f(\mathbf{X}^i) / g(\mathbf{X}^i)}{\sum_{i=1}^k I(\mathbf{X}^i \in \mathcal{A}) f(\mathbf{X}^i) / g(\mathbf{X}^i)} \quad (8.14)$$

Podemos estimar el error cuadrático medio de este estimador mediante el método bootstrap (véase el ejemplo 7e).

Ejemplo 8w Sean X_i variables aleatorias exponenciales independientes con razones respectivas $1/(i+2)$, $i = 1, 2, 3, 4$. Sea $S = \sum_{i=1}^4 X_i$, y suponga que queremos estimar $\theta = E[S | S > 62]$. Para hacerlo, aplicamos el muestreo de importancia con las distribuciones transformadas. Es decir, elegimos un valor t y luego simulamos las X_i con razones $1/(i+2) - t$. Si elegimos $t = 0.14$, entonces $E_t[S] = 68.43$. Así, generamos k conjuntos de variables aleatorias exponenciales X_i con razones $1/(i+2) - 0.14$, $i = 1, 2, 3, 4$, y sea S_j la suma del j -ésimo conjunto, $j = 1, \dots, k$. Entonces podemos estimar

$$E[S I(S > 62)] \quad \text{con} \quad \frac{C}{k} \sum_{j=1}^k S_j I(S_j > 62) e^{-0.14 S_j}$$

$$E[I(S > 62)] \quad \text{con} \quad \frac{C}{k} \sum_{j=1}^k I(S_j > 62) e^{-0.14 S_j}$$

donde $C = \prod_{i=1}^4 \frac{1}{1 - 0.14(i+2)} = 81.635$. El estimador de θ es

$$\hat{\theta} = \frac{\sum_{j=1}^k S_j I(S_j > 62) e^{-0.14S_j}}{\sum_{j=1}^k I(S_j > 62) e^{-0.14S_j}}$$

El método del muestreo de importancia también es útil pues permite estimar dos (o más) cantidades distintas en una única simulación. Por ejemplo, suponga que

$$\theta_1 = E[h(\mathbf{Y})] \quad \text{y} \quad \theta_2 = E[h(\mathbf{W})]$$

donde \mathbf{Y} y \mathbf{W} son vectores aleatorios con funciones de densidad conjunta f y g , respectivamente. Si ahora simulamos \mathbf{W} , podemos utilizar en forma simultánea $h(\mathbf{W})$ y $h(\mathbf{W})f(\mathbf{W})/g(\mathbf{W})$ como estimadores de θ_2 y θ_1 , respectivamente. Por ejemplo, suponga que simulamos T , el tiempo total dentro del sistema para los primeros r clientes en un sistema de línea de espera, en el cual la distribución del servicio es exponencial con media 2. Si ahora decidimos que deberíamos considerar el mismo sistema pero con una distribución de servicio con distribución gamma y parámetros (2, 1), entonces no es necesario repetir la simulación, pues basta tomar el estimador

$$T \frac{\prod_{i=1}^r S_i \exp\{-S_i\}}{\prod_{i=1}^r \left(\frac{1}{2} \exp\{-S_i/2\}\right)} = 2^r T \exp\left\{-\sum_{i=1}^r \frac{S_i}{2}\right\} \prod_{i=1}^r S_i$$

donde S_i es el tiempo de servicio (exponencial) generado para el cliente i . [Lo anterior es consecuencia de que la densidad del tiempo de servicio exponencial es $g(s) = 1/2 e^{-s/2}$, mientras que la densidad gamma (2, 1) es $f(s) = se^{-s}$.]

8.6 Uso de números aleatorios comunes

Suponga que un conjunto de n trabajos deben ser procesados por un par de máquinas idénticas. Sea T_i el tiempo de procesamiento para el trabajo i , $i = 1, \dots, n$. Estamos interesados en comparar el tiempo que tarda en concluir el procesamiento de todos los trabajos según dos diferentes políticas de decisión del orden de procesamiento de los trabajos. Cuando una máquina queda libre, la primera política (que considera primero el trabajo mayor), siempre elige el trabajo restante con mayor tiempo de procesamiento, mientras que la segunda política (que considera primero el trabajo menor),

siempre elige el trabajo con menor tiempo de procesamiento. Por ejemplo, si $n = 3$ y $T_1 = 2$, $T_2 = 5$ y $T_3 = 3$, entonces la política de primero el trabajo mayor concluiría el procesamiento en el instante 5, mientras que la política de primero el menor trabajo no terminaría sino hasta el instante 7. Queremos utilizar la simulación para comparar la diferencia esperada en los tiempos de conclusión según ambas políticas, cuando los tiempos de procesamiento de los trabajos, T_1, \dots, T_n , son variables aleatorias con una distribución dada F .

En otras palabras, si $g(t_1, \dots, t_n)$ es el tiempo que se requiere para procesar los n trabajos, los cuales tienen tiempos de procesamiento t_1, \dots, t_n cuando seguimos la política del mayor trabajo primero y si $h(t_1, \dots, t_n)$ es el tiempo cuando practicamos la política de primero el menor trabajo, entonces estamos interesados en utilizar simulación para estimar

$$\theta = \theta_1 - \theta_2$$

donde

$$\theta_1 = E[g(\mathbf{T})], \quad \theta_2 = E[h(\mathbf{T})], \quad \mathbf{T} = (T_1, \dots, T_n)$$

Si ahora generamos el vector \mathbf{T} para calcular $g(\mathbf{T})$, surge la cuestión de si debemos utilizar estos mismos valores generados para calcular $h(\mathbf{T})$ o si es más eficiente generar un conjunto independiente para estimar θ_2 . Para responder esta cuestión, suponga que tomamos $\mathbf{T}^* = (T_1^*, \dots, T_n^*)$, con la misma distribución que \mathbf{T} , para estimar θ_2 . Entonces la varianza del estimador $g(\mathbf{T}) - h(\mathbf{T}^*)$ de θ es

$$\begin{aligned} \text{Var}(g(\mathbf{T}) - h(\mathbf{T}^*)) &= \text{Var}(g(\mathbf{T})) + \text{Var}(h(\mathbf{T}^*)) - 2 \text{Cov}(g(\mathbf{T}), h(\mathbf{T}^*)) \\ &= \text{Var}(g(\mathbf{T})) + \text{Var}(h(\mathbf{T})) - 2 \text{Cov}(g(\mathbf{T}), h(\mathbf{T}^*)) \end{aligned} \quad (8.15)$$

Por lo tanto, si $g(\mathbf{T})$ y $h(\mathbf{T})$ están correlacionadas en forma positiva (es decir, si su covarianza es positiva), entonces la varianza del estimador de θ es menor si utilizamos el mismo conjunto de valores aleatorios generados \mathbf{T} para calcular $g(\mathbf{T})$ y $h(\mathbf{T})$ que si empleáramos un conjunto independiente \mathbf{T}^* para calcular $h(\mathbf{T}^*)$ [en ese último caso, la covarianza en (8.15) sería igual a 0].

Como g y h son funciones crecientes de sus argumentos y como las funciones crecientes de variables aleatorias independientes están correlacionadas en forma positiva (véase una demostración en el apéndice del capítulo), se tiene que en el caso anterior es más eficiente comparar en forma sucesiva las políticas usando siempre el mismo conjunto de tiempos de trabajo generados para ambas políticas.

Como regla empírica general al comparar diferentes políticas operativas en un ambiente determinado en forma aleatoria, después de simular el estado del ambiente debemos evaluar todas las políticas con este ambiente. Es decir, si el ambiente está determinado mediante el vector \mathbf{T} y $g(\mathbf{T})$ se obtiene de la política i en el estado

ambiental \mathbf{T} , después de simular el valor del vector aleatorio \mathbf{T} uno debe evaluar todos los resultados $g_i(\mathbf{T})$ para ese valor de \mathbf{T} .

Apéndice: verificación del método de las variables antitéticas al estimar el valor esperado de funciones monótonas

El siguiente teorema es la clave para mostrar que el uso de variables antitéticas conduce a una reducción de la varianza, en comparación con la generación de un nuevo conjunto independiente de números aleatorios, siempre que la función h sea monótona en cada una de sus coordenadas.

Teorema Si X_1, \dots, X_n son independientes, entonces para cualesquiera funciones crecientes f y g de n variables

$$E[f(\mathbf{X})g(\mathbf{X})] \geq E[f(\mathbf{X})]E[g(\mathbf{X})] \quad (8.16)$$

donde $\mathbf{X} = (X_1, \dots, X_n)$.

Demostración La demostración es por inducción sobre n . Para demostrarla cuando $n = 1$, sean f y g funciones crecientes de una sola variable. Entonces, para cualesquiera x y y ,

$$[f(x) - f(y)][g(x) - g(y)] \geq 0$$

pues si $x \geq y$ ($x \leq y$), entonces ambos factores son no negativos (no positivos). Por lo tanto, para cualesquiera variables aleatorias X y Y ,

$$[f(X) - f(Y)][g(X) - g(Y)] \geq 0$$

implica que

$$E\{[f(X) - f(Y)][g(X) - g(Y)]\} \geq 0$$

o, en forma equivalente,

$$E[f(X)g(X)] + E[f(Y)g(Y)] \geq E[f(X)g(Y)] + E[f(Y)g(X)]$$

Si ahora suponemos que X y Y son independientes e idénticamente distribuidas, como en este caso,

$$E[f(X)g(X)] = E[f(Y)g(Y)]$$

$$E[f(X)g(Y)] = E[f(Y)g(X)] = E[f(X)]E[g(X)]$$

obtenemos el resultado cuando $n = 1$.

Así, suponga que la ecuación (8.16) se cumple para $n - 1$ variables, y ahora suponga que X_1, \dots, X_n son independientes y que f y g son funciones crecientes. Entonces

$$\begin{aligned} E[f(\mathbf{X})g(\mathbf{X})|X_n = x_n] &= E[f(X_1, \dots, X_{n-1}, x_n)g(X_1, \dots, X_{n-1}, x_n)|X_n = x_n] \\ &= E[f(X_1, \dots, X_{n-1}, x_n)g(X_1, \dots, X_{n-1}, x_n)] \\ &\quad \text{por la independencia} \\ &\geq E[f(X_1, \dots, X_{n-1}, x_n)]E[g(X_1, \dots, X_{n-1}, x_n)] \\ &\quad \text{por la hipótesis de inducción} \\ &= E[f(\mathbf{X})|X_n = x_n]E[g(\mathbf{X})|X_n = x_n] \end{aligned}$$

Por lo tanto,

$$E[f(\mathbf{X})g(\mathbf{X})|X_n] \geq E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]$$

y, al obtener las esperanzas de ambos lados,

$$\begin{aligned} E[f(\mathbf{X})g(\mathbf{X})] &\geq E[E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]] \\ &\geq E[f(\mathbf{X})]E[g(\mathbf{X})] \end{aligned}$$

La última desigualdad se debe a que $E[f(\mathbf{X})|X_n]$ y $E[g(\mathbf{X})|X_n]$ son ambas funciones crecientes de X_n , y entonces, por el resultado para $n = 1$,

$$\begin{aligned} E[E[f(\mathbf{X})|X_n]E[g(\mathbf{X})|X_n]] &\geq E[E[f(\mathbf{X})|X_n]]E[E[g(\mathbf{X})|X_n]] \\ &= E[f(\mathbf{X})]E[g(\mathbf{X})] \end{aligned}$$

Corolario Si $h(x_1, \dots, x_n)$ es una función monótona de cada uno de sus argumentos, entonces, para un conjunto U_1, \dots, U_n de números aleatorios independientes,

$$\text{Cov}[h(U_1, \dots, U_n), h(1 - U_1, \dots, 1 - U_n)] \leq 0$$

Demostración Podemos redefinir h para suponer, sin pérdida de generalidad, que h es creciente en los primeros r argumentos y decreciente en los últimos $n - r$. Por lo tanto, si

$$f(x_1, \dots, x_n) = h(x_1, \dots, x_r, 1 - x_{r+1}, \dots, 1 - x_n)$$

$$g(x_1, \dots, x_n) = -h(1 - x_1, \dots, 1 - x_r, x_{r+1}, \dots, x_n)$$

se tiene que f y g son funciones crecientes. Así, por el teorema anterior,

$$\text{Cov}[f(U_1, \dots, U_n), g(U_1, \dots, U_n)] \geq 0$$

o, en forma equivalente,

$$\text{Cov}[h(U_1, \dots, U_r, 1 - U_{r+1}, \dots, 1 - U_n), h(1 - U_1, \dots, 1 - U_r, U_{r+1}, \dots, U_n)] \leq 0$$

Obtenemos entonces el resultado, pues el vector aleatorio $h(U_1, \dots, U_n)$, $h(1 - U_1, \dots, 1 - U_n)$ tiene la misma distribución conjunta que el vector aleatorio

$$h(U_1, \dots, U_r, 1 - U_{r+1}, \dots, 1 - U_n), h(1 - U_1, \dots, 1 - U_r, U_{r+1}, \dots, U_n) \quad \blacksquare$$

Ejercicios

1. Suponga que queremos estimar θ , donde

$$\theta = \int_0^1 e^{x^2} dx$$

Muestre que generar un número aleatorio U y luego utilizar el estimador $e^{U^2}(1 + e^{1-2U})/2$ es mejor que generar dos números aleatorios U_1 y U_2 y emplear $[\exp(U_1^2) + \exp(U_2^2)]/2$.

2. Explique cómo es posible usar variables antitéticas para obtener una estimación por simulación de la cantidad

$$\theta = \int_0^1 \int_0^1 e^{(x+y)^2} dy dx$$

En este caso, ¿es claro que el uso de las variables antitéticas es mejor que generar un nuevo par de números aleatorios?

3. Sean X_i , $i = 1, \dots, 5$, variables aleatorias exponenciales independientes, cada una con media 1, y considere la cantidad θ definida por

$$\theta = P\left\{\sum_{i=1}^5 iX_i \geq 21.6\right\}$$

- (a) Explique cómo podemos utilizar la simulación para estimar θ .
 (b) Dé el estimador mediante variables antitéticas.
 (c) ¿En este caso es eficiente el uso de variables antitéticas?
4. Muestre que si X y Y tienen la misma distribución, entonces $\text{Var}[(X + Y)/2] \leq \text{Var}(X)$ y concluya que el uso de variables antitéticas nunca puede incrementar la varianza (aunque no necesariamente es tan eficiente como generar un conjunto independiente de números aleatorios).

5. (a) Si Z es una variable aleatoria normal estándar, diseñe un estudio con variables antitéticas para estimar $\theta = E[Z^3 e^Z]$.
 (b) A partir de lo anterior realice la simulación para obtener un intervalo de longitud no mayor que 0.1 donde pueda afirmar, con 95 por ciento de confianza, que contiene al valor de θ .
6. Suponga que X es una variable aleatoria exponencial con media 1. Dé otra variable aleatoria que esté correlacionada en forma negativa con X y que también sea exponencial con media 1.
7. Verifique la ecuación (8.1).
8. Verifique la ecuación (8.2).
9. Explique por qué no es usual obtener un gran incremento en la eficiencia al utilizar las variables de control y las variables antitéticas, en vez de emplear sólo unas variables o las otras.
10. Sea U_n , $n \geq 1$, una sucesión de variables aleatorias independientes uniformes en $(0, 1)$. Defina

$$S = \min(n: U_1 + \dots + U_n > 1)$$

Puede demostrarse que S tiene la misma distribución que N , del ejemplo 3e, y de modo que $E[S] = e$. Además, si hacemos

$$T = \min(n: 1 - U_1 + \dots + 1 - U_n > 1)$$

entonces se puede mostrar que $S + T$ tiene la misma distribución que $N + M$ del ejemplo 3e. Esto sugiere el uso de $(S + T + N + M)/4$ para estimar a e . Utilice simulación para estimar $\text{Var}(N + M + S + T)/4$.

11. En ciertas situaciones, se simula una variable aleatoria X con media conocida para obtener una estimación de $P\{X \leq a\}$ para una constante dada a . El estimador de simulación en bruto a partir de una sola ejecución es I , donde

$$I = \begin{cases} 1 & \text{si } X \leq a \\ 0 & \text{si } X > a \end{cases}$$

Como es claro que I y X están correlacionados en forma negativa, un intento natural de reducción de la varianza consiste en usar X como control y por ende utilizar un estimador de la forma $I + c(X - E[X])$.

- (a) Determine el porcentaje posible de reducción de la varianza sobre el estimador en bruto I (utilizando la mejor c) si X fuese uniforme en $(0, 1)$.
 (b) Repita (a) si X fuese exponencial con media 1.
 (c) Explique por qué sabíamos que I y X estaban correlacionados en forma negativa.
12. Muestre que $\text{Var}(\alpha X + (1 - \alpha)W)$ se minimiza con α igual al valor dado en la ecuación (8.3), y determine la varianza resultante.

13. (a) Explique la forma de utilizar las variables de control para estimar θ en el ejercicio 1.
 (b) Realice 100 ejecuciones de simulación, con el control dado en (a), para estimar primero c^* y luego la varianza del estimador.
 (c) Utilice los mismos datos de la parte (b) y determine la varianza del estimador mediante variables antitéticas.
 (d) ¿Cuál de los dos tipos de técnicas de reducción de varianza fue mejor en este ejemplo?
14. Repita el ejercicio 13 para θ dada en el ejercicio 2.
15. Repita el ejercicio 13 para θ dada en el ejercicio 3.
16. Muestre que al estimar $\theta = E[(1 - U^2)^{1/2}]$ es mejor utilizar U^2 que U como variable de control. Para ello, utilice simulación para aproximar las covarianzas necesarias.
17. Cinco elementos, numerados 1, 2, 3, 4, 5, se ordenan inicialmente en forma aleatoria (es decir, el orden inicial es una permutación aleatoria de 1, 2, 3, 4, 5). En cada etapa, se elige uno de los elementos y se coloca al frente de la lista. Es decir, si el orden actual es 2, 3, 4, 1, 5 y se elige el elemento 1, entonces el nuevo orden es 1, 2, 3, 4, 5. Suponga que cada selección es, de manera independiente, el elemento i con probabilidad p_i , donde $p_1 = \frac{1}{15}$, $p_2 = \frac{2}{15}$, $p_3 = \frac{3}{15}$, $p_4 = \frac{4}{15}$, $p_5 = \frac{5}{15}$. Sea L_j la posición del j -ésimo elemento por seleccionar y sea $L = \sum_{j=1}^{100} L_j$. Estamos interesados en utilizar simulación para estimar $E[L]$.
- (a) Explique cómo podríamos utilizar simulación para estimar $E[L]$.
 (b) Calcule $E[N_i]$, donde N_i es el número de veces que se elige el elemento i en las 100 selecciones.
 (c) Sea $Y = \sum_{i=1}^5 iN_i$. ¿Cree que Y está correlacionada en forma positiva o negativa con L ?
 (d) Desarrolle un estudio para estimar L , con Y como variable de control.
 (e) Dé un método distinto, con la idea del ejemplo 8i y desarrolle un estudio para determinar la eficiencia de este método.
18. Sean X y Y independientes, con distribuciones respectivas F y G y con valores esperados μ_x y μ_y . Para un valor dado t , estamos interesados en estimar $\theta = P\{X + Y \leq t\}$.
- (a) Dé el método de simulación en bruto para estimar θ .
 (b) Utilice "condicionamiento" para obtener un mejor estimador.
 (c) Dé una variable de control que se pueda utilizar para mejorar aún más el estimador de (b).
19. (Método dentro-fuera). Sea g una función acotada sobre el intervalo $[0, 1]$ (por ejemplo, suponga que $0 \leq g(x) \leq b$ siempre que $0 \leq x \leq 1$) y suponga que estamos interesados en utilizar simulación para aproximar $\theta = \int_0^1 g(x) dx$. El método dentro-

fuera para lograrlo consiste en generar un par de números aleatorios independientes U_1 y U_2 . Sean $X = U_1$, $Y = bU_2$ tales que el punto aleatorio (X, Y) se distribuye de manera uniforme en un rectángulo de longitud 1 y altura b . Sea

$$I = \begin{cases} 1 & \text{si } Y < g(x) \\ 0 & \text{en caso contrario} \end{cases}$$

Es decir, I es igual a 1 si el punto aleatorio (X, Y) está dentro del área sombreada de la figura 8.4.

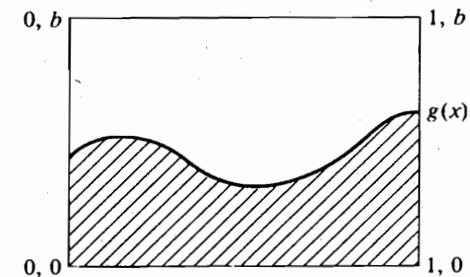


Figura 8.4

- (a) Muestre que $E[I] = [\int_0^1 g(x) dx]/b$.
 (b) Muestre que $\text{Var}(bI) \geq \text{Var}(g(U))$, por lo cual el estimador del método dentro-fuera tiene una mayor varianza que el cálculo de g de un número aleatorio.
20. Suponga que X es exponencial con media 1; y dado $X = x$, Y es exponencial con media x (así, X y Y son variables aleatorias dependientes). Dé una forma eficiente de estimar $P\{XY \leq 3\}$.
21. Sean X y Y exponenciales independientes, donde X tiene media 1 y Y tiene media 2, y suponga que queremos utilizar simulación para estimar $P\{X + Y > 4\}$. Si fuésemos a utilizar la esperanza condicional para reducir la varianza del estimador, ¿condicionaría sobre X o sobre Y ? Explique su razonamiento.
22. Sean X y Y variables aleatorias normales independientes, ambas con media 1 y varianza 1, y sea $\theta = E[e^{XY}]$.
- (a) Explique el método de simulación para estimar θ .
 (b) Dé una variable de control y explique la forma de emplearla para obtener un estimador con una varianza menor que la del estimador de simulación en bruto de la parte (a).
 (c) Dé otra variable de control que intuitivamente sea mejor que la dada en la parte (b). *Sugerencia:* recuerde el desarrollo en serie de $f(x) = e^x$.
 (d) Suponga que ha generado X y Y . ¿Cuál sería el estimador de θ mediante las variables antitéticas?

- (e) ¿Tendría necesariamente el estimador de la parte (d) una menor varianza que el estimador de simulación en bruto basado en dos pares de valores X y Y ? ¿Por qué?
- (f) Utilice la esperanza condicional para mejorar el estimador de simulación en bruto. *Sugerencia:* si W es normal con media μ y varianza σ^2 , entonces $E[e^W] = e^{\mu + \sigma^2/2}$.
- (g) Mejore el estimador de la parte (f) mediante una variable de control.
- (h) Estime θ de manera eficiente. Detenga la simulación cuando esté 95 por ciento seguro de que su estimación es correcta hasta un 0.1.
23. Suponga que los clientes llegan de acuerdo con un proceso Poisson con razón λ , a la línea de espera de una estación con un único servidor. Al llegar, los clientes pasan a servicio, si el servidor está libre, o bien se forman en la fila. Al terminar un servicio, el primer cliente de la fila, si hay clientes formados, entra a servicio. Todos los tiempos de servicio son variables aleatorias independientes con distribución G . Suponga que se ha planeado que el servidor tome un descanso en el tiempo T si el sistema está vacío en ese momento o en el primer momento posterior a T cuando el sistema esté vacío. Sea X la cantidad de tiempo después de T en que el servidor toma un descanso y suponga que queremos valernos de la simulación para estimar $E[X]$. Explique la forma de utilizar la esperanza condicional para obtener un estimador eficiente de $E[X]$.
- Sugerencia:* considere la situación en el instante T relacionada con el tiempo de servicio restante del cliente que actualmente está en el servicio y el número de clientes en la fila (este problema requiere cierto conocimiento de la teoría del periodo de ocupación $M/G/1$).
24. Considere una cola con un único servidor, cuyos clientes llegan de acuerdo con un proceso Poisson con razón 2 por minuto y los tiempos de servicio se distribuyen en forma exponencial con media de 1 minuto. Sea T_i el tiempo que el cliente i pasa dentro del sistema. Nos interesa utilizar simulación para estimar $\theta = E[T_1 + \dots + T_{10}]$.
- (a) Realice una simulación para estimar la varianza del estimador de simulación en bruto. Es decir, estime $\text{Var}(T_1 + \dots + T_{10})$.
- (b) Haga una simulación para determinar la mejora sobre el estimador en bruto obtenida empleando variables antitéticas.
- (c) Realice una simulación para determinar la mejora sobre el estimador en bruto obtenido, utilizando $\sum_{i=1}^{10} S_i$ como variable de control, donde S_i es el i -ésimo tiempo de servicio.
- (d) Realice una simulación para determinar la mejora sobre el estimador en bruto obtenido; use $\sum_{i=1}^{10} S_i - \sum_{i=1}^9 I_i$ como variable de control, donde I_i es el tiempo entre las i -ésima e $(i+1)$ -ésima llegadas.
- (e) Realice una simulación para determinar la mejora sobre el estimador en bruto

- obtenido utilizando el estimador $\sum_{i=1}^{N_0} E[T_i | N_i]$, donde N_i es el número de clientes en el sistema cuando llega el cliente i (de modo que $N_1 = 0$).
25. En el ejemplo 8q, calcule $E[X | i]$ para $i = 0, 1, 2$.
26. Estime la varianza del estimador de simulación en bruto del pago esperado en el modelo de video póquer del ejemplo 8q. Luego, estime la varianza mediante la reducción de varianza sugerida en ese ejemplo. ¿Cuál es su estimación del pago esperado?, (si es menor que 1, entonces el juego no es justo para el jugador).
27. Considere un sistema de 20 componentes independientes, donde el componente i deja de ser funcional con probabilidad $0.5 + i/50$, $i = 1, \dots, 20$. Sea X el número de componentes no funcionales. Utilice simulación para estimar de manera eficiente $P\{X \leq 5\}$.
28. Estime $P\{X = 5 | X \leq 5\}$ en el ejercicio anterior.
29. Si X es tal que $P\{0 \leq X \leq a\} = 1$, muestre que
- (a) $E[X^2] \leq aE[X]$.
- (b) $\text{Var}(X) \leq E[X](a - E[X])$.
- (c) $\text{Var}(X) \leq a^2/4$.
- Sugerencia:* recuerde que $\max_{0 \leq p \leq 1} p(1-p) = \frac{1}{4}$.
30. En el ejemplo 8w, dé una cota superior analítica sobre $P\{S > 62\}$.
31. Utilice simulación para estimar $E[S | S > 200]$ en el ejemplo 8w.
32. Suponga que tenemos una "caja negra" que puede generar el valor de una variable aleatoria gamma con parámetros $\frac{3}{2}$ y 1. Explique la forma de utilizar esta caja negra para aproximar $E[e^X/(X+1)]$, donde X es una variable aleatoria exponencial con media 1.
33. En el ejercicio 13 del capítulo 6, suponga que estamos interesados en utilizar simulación para estimar p , la probabilidad de que el sistema falle hasta cierto instante fijo t . Si p es muy pequeña, explique la forma de aprovechar el muestreo de importancia para obtener un estimador más eficiente que el estimador en bruto. Elija algunos valores para α , C y t que hagan p pequeña y realice una simulación para estimar la varianza de un estimador de muestreo de importancia, así como el estimador de simulación en bruto de p .
34. Considere dos métodos distintos para fabricar un producto. La ganancia en ambos depende del valor de un parámetro α , y sea $v_i(\alpha)$ la ganancia del método i como función de α . Suponga que el método 1 funciona mejor para valores pequeños de α , en los cuales $v_1(\alpha)$ es una función decreciente de α , mientras que el método 2 funciona mejor para valores grandes de α , en los cuales $v_2(\alpha)$ es una función creciente de α . Si el valor diario de α es una variable aleatoria que proviene de la distribución F , al comparar la ganancia promedio de estos dos métodos, ¿debemos generar un único valor de α y calcular

las ganancias para esta α , o debemos generar α_1 y α_2 y luego calcular $v_i(\alpha_i)$, $i = 1, 2$?

35. Considere una lista de n nombres en la que n es muy grande y suponga que un nombre dado puede aparecer muchas veces en la lista. Sea $N(i)$ el número de veces que el nombre de la posición i aparece en la lista, $i = 1, \dots, n$, y sea D el número de nombres distintos en la lista. Estamos interesados en utilizar la simulación para estimar $\theta = E[D]$.
- (a) Justifique que $D = \sum_{i=1}^n \frac{1}{N(i)}$.
Supongamos que X tiene la misma probabilidad de tomar los valores $1, \dots, n$. Determine el nombre en la posición X y recorra la lista desde el principio, deteniéndose al encontrar ese nombre. Sea $Y = 1$ si el nombre aparece por vez primera en la posición X y $Y = 0$ en caso contrario (es decir, $Y = 1$ si la primera aparición del nombre es en la posición X).
 - (b) Justifique que $E[Y|N(X)] = \frac{1}{N(X)}$.
 - (c) Justifique que $E[nY] = \theta$.
 - (d) Ahora, sea $W = 1$ si la posición X es la última vez que el nombre en esa posición aparece en la lista, y sea $W = 0$ en caso contrario (es decir, $W = 1$ si al ir de atrás hacia adelante en la lista, el nombre aparece por vez primera en la posición X). Justifique que $n(W + Y)/2$ es un estimador insesgado de θ .
 - (e) Justifique que si cada nombre de la lista aparece al menos dos veces, entonces el estimador de la parte (d) es un mejor estimador de θ que $(nY_1 + nY_2)/2$, donde Y_1 y Y_2 son independientes y distribuidas como Y .
 - (f) Justifique que $n/(N(X))$ tiene menor varianza que el estimador de la parte (e), aunque el estimador de la parte (e) puede seguir siendo más eficaz cuando hay demasiadas copias, pues su proceso de búsqueda es más rápido.

Bibliografía

- Hammersley, J. M. y D. C. Handscomb, *Monte Carlo Methods*, Nueva York, Wiley, 1964.
- Hammersley, J. M. y K. W. Morton, "A New Monte Carlo Technique: Antithetic Variables", en *Proc. Cambridge Phil. Soc.*, **52**, 1956, pp. 449-474.
- Lavenberg, S. S. y P. D. Welch, "A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations", en *Management Science*, **27**, 1981, pp. 322-335.
- Morgan, B. J. T., *Elements of Simulation*, Londres, Chapman and Hall, 1983.
- Ripley, B., *Stochastic Simulation*, Nueva York, Wiley, 1986.
- Rubinstein, R. Y., *Simulation and the Monte Carlo Method*, Nueva York, Wiley, 1981.
- Siegmund, D., "Importance Sampling in the Monte Carlo Study of Sequential Tests", en *Annals of Statistics*, **4**, 1976, pp. 673-684.

CAPÍTULO

9

Técnicas de validación estadística

Introducción

En este capítulo consideramos algunos procedimientos estadísticos útiles para dar validez a los modelos de simulación. Las secciones 9.1 y 9.2 consideran las pruebas de bondad de ajuste, que son útiles para saber si una distribución de probabilidad supuesta es congruente con un conjunto dado de datos. En la sección 9.1 suponemos que la distribución supuesta está totalmente especificada, mientras que en la sección 9.2 suponemos que está especificada salvo ciertos parámetros; por ejemplo, puede ser Poisson con una media desconocida. En la sección 9.3 mostramos la forma de verificar la hipótesis de que dos muestras separadas de datos provienen de la misma población subyacente, como ocurriría con datos reales y simulados cuando el modelo matemático supuesto es una representación precisa de la realidad. Los resultados de la sección 9.3 son de particular ayuda para verificar la validez de un modelo de simulación. En esta sección también se presenta una generalización del caso de varias muestras. Por último, en la sección 9.4 mostramos la forma de acudir a datos reales para verificar la hipótesis de que el proceso generador de los datos es un proceso Poisson no homogéneo. El caso de un proceso Poisson homogéneo también se analiza en esta sección.

9.1 Pruebas de bondad de ajuste

Con frecuencia, uno comienza un análisis probabilístico de un fenómeno dado estableciendo como hipótesis que algunos de sus elementos aleatorios tienen una distribución de probabilidad particular. Por ejemplo, podríamos iniciar el análisis de una red

de tráfico suponiendo que el número diario de accidentes tiene una distribución Poisson. Para verificar tales hipótesis en forma estadística se observan los datos y se decide después si la hipótesis de una distribución de probabilidad particular es congruente con éstos. Las pruebas estadísticas se llaman *pruebas de bondad de ajuste*.

Una forma de realizar una prueba de bondad de ajuste es separar primero los valores posibles de una cantidad aleatoria en un número finito de regiones. Luego se observa una muestra de valores de esta cantidad y se compara entre el número de ellos que caen en cada una de las regiones y los números teóricos esperados cuando la distribución de probabilidad especificada realmente controla los datos.

En esta sección analizamos las pruebas de bondad de ajuste en las que todos los parámetros de la supuesta distribución están especificados; en la siguiente sección consideramos las pruebas cuando algunos de los parámetros no están dados. Primero consideramos el caso de una distribución discreta y luego el de una distribución continua.

PRUEBA DE BONDAD DE AJUSTE JI-CUADRADA PARA DATOS DISCRETOS

Suponga que vamos a observar n variables aleatorias independientes Y_1, \dots, Y_n , cada una de las cuales toma uno de los valores $1, 2, \dots, k$ y que estamos interesados en verificar la hipótesis de que $\{p_i, i = 1, \dots, k\}$ es la función de masa de probabilidad de estas variables aleatorias. Es decir, si Y representa cualquiera de las Y_j , la hipótesis por verificar, que denotamos por H_0 y llamamos la *hipótesis nula*, es

$$H_0: P\{Y = i\} = p_i \quad i = 1, \dots, k$$

Para verificar la hipótesis anterior, sea $N_i, i = 1, \dots, k$ el número de Y_j que son iguales a i . Como cada Y_j , de manera independiente, es igual a i con probabilidad $P\{Y = i\}$, se tiene que según H_0 , N_i es binomial con parámetros n y p_i . Por lo tanto, cuando H_0 es cierta,

$$E[N_i] = np_i$$

de modo que $(N_i - np_i)^2$ es un indicador de la probabilidad de que p_i sea igual a la probabilidad de que $Y = i$. Cuando esto es grande, digamos en relación con np_i , entonces es una indicación de que H_0 es incorrecta. De hecho, tal razonamiento nos conduce a considerar la cantidad

$$T = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

y rechazar la hipótesis nula cuando T es grande.

Mientras los valores pequeños de la cantidad de prueba T son una evidencia en favor de la hipótesis H_0 , los valores grandes indican su falsedad. Suponga que los

datos verdaderos hacen que la cantidad de prueba T asuma el valor t . Para ver lo poco probable que sería un resultado grande si la hipótesis nula fuese cierta, definimos el valor p , como

$$\text{valor } p = P_{H_0}\{T \geq t\}$$

donde hemos usado la notación P_{H_0} para indicar que la probabilidad debe calcularse según la hipótesis de que H_0 es correcta. Por lo tanto, el valor p proporciona la probabilidad de que ocurra un valor grande de T , como el observado, si la hipótesis nula fuese cierta. Es usual rechazar la hipótesis nula diciendo que parece no ser congruente con los datos, cuando se obtiene un valor p pequeño (por lo general, un valor menor que 0.05, o de manera más conservadora, 0.01, se considera crítico) y, en caso contrario, aceptar la hipótesis nula, afirmando que parece ser congruente con los datos.

Después de observar el valor (digamos t) de la cantidad de prueba, resta determinar la probabilidad

$$\text{valor } p = P_{H_0}\{T \geq t\}$$

Una aproximación razonablemente buena de esta probabilidad se puede obtener mediante el resultado clásico de que, para valores grandes de n , T tiene aproximadamente una distribución ji-cuadrada con $k - 1$ grados de libertad cuando la H_0 es verdadera. Por lo tanto,

$$\text{valor } p \approx P\{X_{k-1}^2 \geq t\} \quad (9.1)$$

donde X_{k-1}^2 es una variable aleatoria ji-cuadrada con $k - 1$ grados de libertad. La probabilidad ji-cuadrada se puede obtener ejecutando el programa 9-1 del apéndice.

Ejemplo 9a Considere una cantidad aleatoria que puede tomar cualquiera de los valores 1, 2, 3, 4, 5 y suponga que queremos verificar la hipótesis de que estos valores tienen la misma probabilidad. Es decir, queremos verificar

$$H_0: p_i = 0.2 \quad i = 1, \dots, 5$$

Si una muestra de tamaño 50 arrojó los siguientes valores de N_i :

$$12, 5, 19, 7, 7$$

entonces el valor p aproximado se obtiene de la manera siguiente: El valor del estadístico de prueba T está dado por

$$T = \frac{4 + 25 + 81 + 9 + 9}{10} = 12.8$$

La ejecución del programa 9-1 indica que

$$\text{valor } p \approx P\{X_4^2 > 12.8\} = 0.0122$$

Para tal valor p tan bajo, la hipótesis de que todos los resultados son igualmente probables sería rechazada. ■

Si la aproximación al valor p dada por la ecuación (9.1) no es demasiado pequeña, digamos del orden de 0.15 o mayor, entonces es evidente que la hipótesis nula no será rechazada, de modo que no hay necesidad de una mejor aproximación. Sin embargo, cuando el valor p está más cerca de un valor crítico (como 0.05 o 0.01), es probable que queramos una estimación más precisa de su valor que la dada por la distribución aproximada ji-cuadrada. Por fortuna, se puede obtener un estimador más preciso mediante un estudio de simulación.

El método de simulación para estimar el valor p del resultado $T = t$ es el siguiente. Para determinar la probabilidad de que T fuese al menos tan grande como t cuando H_0 es cierta, generamos n variables aleatorias independientes $Y_1^{(1)}, \dots, Y_n^{(1)}$, cada una con la función de masa de probabilidad $\{p_i, i = 1, \dots, k\}$; es decir,

$$P\{Y_j^{(1)} = i\} = p_i \quad i = 1, \dots, k, \quad j = 1, \dots, n$$

Ahora, sea

$$N_i^{(1)} = \text{número } j: Y_j^{(1)} = i$$

y sea

$$T^{(1)} = \sum_{i=1}^k \frac{(N_i^{(1)} - np_i)^2}{np_i}$$

Ahora repetimos este procedimiento simulando un segundo conjunto, independiente del primero, de n variables aleatorias independientes $Y_1^{(2)}, \dots, Y_n^{(2)}$ cada una de las cuales tiene la función de masa de probabilidad $\{p_i, i = 1, \dots, k\}$ y luego, como en el caso del primer conjunto, determinar $T^{(2)}$. Al repetir esto un gran número de veces, digamos r , produce r variables aleatorias independientes $T^{(1)}, T^{(2)}, \dots, T^{(r)}$, cada una de las cuales tiene la misma distribución que el estadístico de prueba cuando H_0 es verdadera. Por lo tanto, por la ley de los grandes números, la proporción de T_i que son tan grandes como t será casi igual a la probabilidad de que T sea tan grande como t cuando H_0 sea verdadera; es decir,

$$\frac{\text{número } i: T^{(i)} \geq t}{r} \approx P_{H_0}\{T \geq t\}$$

El programa 9-2 del apéndice utiliza simulación para aproximar el valor p . El usuario debe proporcionar las probabilidades $p_i, i = 1, \dots, k$, el tamaño de la muestra n , el valor de la cantidad de prueba observada T y el número de ejecuciones de simulación deseadas.

Ejemplo 9b Consideremos el ejemplo 9a; esta vez realizaremos una simulación para aproximar el valor p . Para esto ejecutamos el programa 9-2.

```

RUN
ESTE PROGRAMA UTILIZA SIMULACIÓN PARA APROXIMAR EL VALOR
p EN LA PRUEBA DE BONDAD DE AJUSTE
Random number seed (-32768 a 32767)? 6867
PROPORCIONE EL NÚMERO DE POSIBLES VALORES
? 5
PROPORCIONE LAS PROBABILIDADES UNA A LA VEZ
? .2
? .2
? .2
? .2
? .2
PROPORCIONE EL TAMAÑO DE LA MUESTRA
?50
PROPORCIONE LA CANTIDAD DESEADA DE EJECUCIONES DE SIMULACIONES
? 1000
PROPORCIONE EL VALOR DEL ESTADÍSTICO DE PRUEBA
?12.8
LA ESTIMACIÓN DEL VALOR p ES .011
OK

```

PRUEBA DE KOLMOGOROV-SMIRNOV PARA DATOS CONTINUOS

Ahora consideremos la situación en la que Y_1, \dots, Y_n son variables aleatorias independientes y estamos interesados en verificar la hipótesis nula H_0 de que todas tienen una función de distribución común F , donde F es una función de distribución continua dada. Un método para verificar H_0 consiste en separar al conjunto de valores posibles de las Y_j en k intervalos distintos, digamos

$$(y_0, y_1), (y_1, y_2), \dots, (y_{k-1}, y_k), \quad \text{donde } y_0 = -\infty, y_k = +\infty$$

y luego considerar las variables aleatorias discretizadas $Y_j^d, j = 1, \dots, n$, definidas por

$$Y_j^d = i \quad \text{si } Y_j \text{ está en el intervalo } (y_{i-1}, y_i)$$

La hipótesis nula implica que

$$P\{Y_j^d = i\} = F(y_i) - F(y_{i-1}), \quad i = 1, \dots, k$$

lo cual puede verificarse mediante la prueba de bondad de ajuste ji-cuadrada ya presentada.

Sin embargo, existe otra forma de verificar que las Y_j provienen de la función de distribución continua F , que por lo general es más eficaz que la discretización; funciona de la manera siguiente. Después de observar a Y_1, \dots, Y_n , sea F_e la función de distribución empírica definida como

$$F_e(x) = \frac{\#i: Y_i \leq x}{n}$$

Es decir, $F_e(x)$ es la proporción de los valores observados que son menores o iguales a x . Como $F_e(x)$ es un estimador natural de la probabilidad de que una observación sea menor o igual a x , se tiene que si es correcta la hipótesis nula de que F es la distribución subyacente, deberá estar cerca de $F(x)$. Como esto es así para toda x , una cantidad natural sobre la cual basar una prueba para H_0 es la cantidad de prueba

$$D \equiv \text{Máximo}_x |F_e(x) - F(x)|$$

donde el máximo (el lector con más conocimientos matemáticos reconocerá que, desde el punto de vista técnico, deberíamos escribir supremo en vez de máximo) se toma sobre todos los valores de x de $-\infty$ hasta $+\infty$. La cantidad D se llama *estadístico de prueba de Kolmogorov-Smirnov*.

Para calcular el valor de D para un conjunto dado de datos $Y_j = y_j, j = 1, \dots, n$, sean $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ los valores de las y_j en orden creciente. Es decir,

$$y_{(j)} = j\text{-ésimo más pequeño de } y_1, \dots, y_n$$

Por ejemplo, si $n = 3$ y $y_1 = 3, y_2 = 5, y_3 = 1$, entonces $y_{(1)} = 1, y_{(2)} = 3, y_{(3)} = 5$. Como $F_e(x)$ se puede escribir

$$F_e(x) = \begin{cases} 0 & \text{si } x < y_{(1)} \\ \frac{1}{n} & \text{si } y_{(1)} \leq x < y_{(2)} \\ \vdots & \\ \frac{j}{n} & \text{si } y_{(j)} \leq x < y_{(j+1)} \\ \vdots & \\ 1 & \text{si } y_{(n)} \leq x \end{cases}$$

vemos que $F_e(x)$ es constante dentro de los intervalos $(y_{(j-1)}, y_{(j)})$ y luego da un salto de $1/n$ en los puntos $y_{(1)}, \dots, y_{(n)}$. Como $F(x)$ es una función creciente de x acotada por 1,

el valor máximo de $F_e(x) - F(x)$ es no negativo y ocurre en alguno de los puntos $y_{(j)}, j = 1, \dots, n$ (véase la figura 9.1). Es decir,

$$\text{Máximo}_x \{F_e(x) - F(x)\} = \text{Máximo}_{j=1, \dots, n} \left\{ \frac{j}{n} - F(y_{(j)}) \right\} \quad (9.2)$$

De manera análoga, el valor máximo (supremo) de $F(x) - F_e(x)$ es no negativo y ocurre inmediatamente antes de uno de los puntos de salto $y_{(j)}$, de modo que

$$\text{Máximo}_x \{F(x) - F_e(x)\} = \text{Máximo}_{j=1, \dots, n} \left\{ F(y_{(j)}) - \frac{(j-1)}{n} \right\} \quad (9.3)$$

Las ecuaciones (9.2) y (9.3) implican que

$$\begin{aligned} D &= \text{Máximo}_x |F_e(x) - F(x)| \\ &= \text{Máximo} \{ \text{Máximo} \{F_e(x) - F(x)\}, \text{Máximo} \{F(x) - F_e(x)\} \} \\ &= \text{Máximo} \left\{ \frac{j}{n} - F(y_{(j)}), F(y_{(j)}) - \frac{(j-1)}{n}, j = 1, \dots, n \right\} \end{aligned} \quad (9.4)$$

La ecuación (9.4) se puede utilizar para calcular el valor de D .

Suponga ahora que se observan las Y_j y que sus valores son tales que $D = d$.

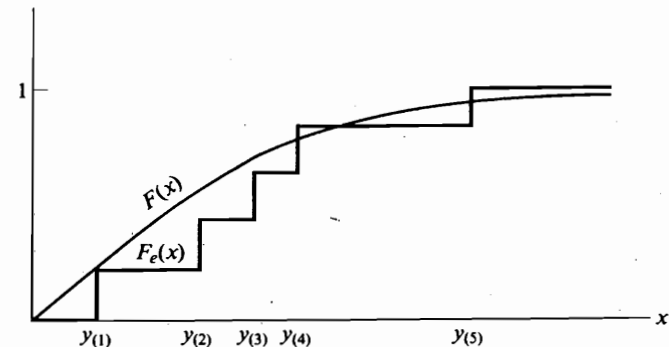


Figura 9.1 $n = 5$

Como un valor grande de D sería incongruente con la hipótesis nula de que F sea la distribución subyacente, tenemos que el valor p para este conjunto de datos está dado por

$$\text{valor } p = P_F\{D \geq d\}$$

donde hemos escrito P_F para hacer explícito que esta probabilidad se debe calcular suponiendo que H_0 es correcta (de modo que F es la distribución subyacente).

El valor p anterior se puede aproximar mediante una simulación facilitada por la siguiente proposición, la cual muestra que $P_F\{D \geq d\}$ no depende de la distribución subyacente F . Esto nos permite estimar el valor p realizando la simulación con cualquier distribución continua F [lo cual nos permite el uso de la distribución uniforme en $(0, 1)$].

Proposición $P_F\{D \geq d\}$ es lo mismo para cualquier distribución continua F .

Demostración

$$\begin{aligned} P_F\{D \geq d\} &= P_F\left\{\text{Máximo}_x \left| \frac{\#i: Y_i \leq x}{n} - F(x) \right| \geq d\right\} \\ &= P_F\left\{\text{Máximo}_x \left| \frac{\#i: F(Y_i) \leq F(x)}{n} - F(x) \right| \geq d\right\} \\ &= P\left\{\text{Máximo}_x \left| \frac{\#i: U_i \leq F(x)}{n} - F(x) \right| \geq d\right\} \end{aligned}$$

donde U_1, \dots, U_n son variables aleatorias independientes, uniformes en $(0, 1)$; la primera igualdad se debe a que F es una función creciente, de modo que $Y \leq x$ es equivalente a $F(Y) \leq F(x)$, y la segunda se debe al resultado (cuya demostración se deja como ejercicio) de que si Y tiene la distribución continua F , entonces la variable aleatoria $F(Y)$ es uniforme en $(0, 1)$.

Continuando con lo anterior, hacemos $y = F(x)$ y observamos que cuando x varía de $-\infty$ a $+\infty$, $F(x)$ varía de 0 a 1, para obtener

$$P_F\{D \geq d\} = P\left\{\text{Máximo}_{0 \leq y \leq 1} \left| \frac{\#i: U_i \leq y}{n} - y \right| \geq d\right\}$$

lo cual muestra que la distribución de D no depende de la distribución real F , cuando H_0 es cierta. ■

La proposición anterior muestra que después de determinar el valor de D mediante los datos, digamos $D = d$, el valor p se puede obtener mediante una simulación con la distribución uniforme en $(0, 1)$. Es decir, generamos un conjunto de n números aleatorios U_1, \dots, U_n y luego verificamos si es cierta o no la desigualdad

$$\text{Máximo}_{0 \leq y \leq 1} \left| \frac{\#i: U_i \leq y}{n} - y \right| \geq d \quad (9.5)$$

Luego, esto se repite muchas veces y la proporción de veces que es válida resulta ser nuestra estimación del valor p del conjunto de datos. Como ya hemos observado, para calcular el lado izquierdo de la desigualdad (9.5) se ordenan los números aleatorios y se emplea después la identidad

$$\begin{aligned} \text{Máx} \left| \frac{\#i: U_i \leq y}{n} - y \right| \\ = \text{Máx} \left\{ \frac{j}{n} - U_{(j)}, U_{(j)} - \frac{(j-1)}{n}, \quad j = 1, \dots, n \right\} \end{aligned}$$

donde $U_{(j)}$ es el j -ésimo valor más pequeño de U_1, \dots, U_n . Por ejemplo, si $n = 3$ y $U_1 = 0.7, U_2 = 0.6, U_3 = 0.4$, entonces $U_{(1)} = 0.4, U_{(2)} = 0.6, U_{(3)} = 0.7$ y el valor de D para este conjunto de datos es

$$D = \text{Máx} \left\{ \frac{1}{3} - 0.4, \frac{2}{3} - 0.6, 1 - 0.7, 0.4, 0.6 - \frac{1}{3}, 0.7 - \frac{2}{3} \right\} = 0.4$$

El programa 9-3 utiliza la simulación para aproximar el valor p de la cantidad D de la prueba de Kolmogorov-Smirnov. El usuario debe proporcionar n , el tamaño del conjunto de datos; d , el valor de D con base en los datos; y el número necesario de ejecuciones de la simulación. El programa genera los valores ordenados de un conjunto de números aleatorios mediante el método bosquejado en el ejercicio 8.

Ejemplo 9c Suponga que queremos verificar la hipótesis de que cierta distribución poblacional es exponencial con media 100; es decir, $F(x) = 1 - e^{-x/100}$. Si los valores (ordenados) de una muestra de tamaño 10 de esta distribución son

66, 72, 81, 94, 112, 116, 124, 140, 145, 155

¿qué conclusión podemos extraer?

Para responder, primero tomamos la ecuación (9.4) para calcular el valor de la cantidad D de la prueba de Kolmogorov-Smirnov. Después de algunos cálculos, obte-

nemos el resultado $D = 0.4831487$. Para obtener una aproximación del valor p , ejecutamos el programa 9-3, el cual da la siguiente salida:

```

RUN
ESTE PROGRAMA UTILIZA SIMULACIÓN PARA APROXIMAR EL VALOR P
DE LA PRUEBA DE KOLMOGOROV-SMIRNOV
Random number seed (-32768 a 32767)? 4567
PROPORCIONE EL VALOR DE LA CANTIDAD DE PRUEBA
? 0.4831487
PROPORCIONE EL TAMAÑO DE LA MUESTRA
? 10
PROPORCIONE EL NÚMERO DESEADO DE EJECUCIONES DE SIMULACIÓN
? 500
EL VALOR p APROXIMADO ES 0.012
OK

```

Como el valor p es muy bajo (es muy poco probable que el menor de un conjunto de 10 valores con la distribución exponencial y media 100 sea tan grande como 66), la hipótesis sería rechazada. ■

9.2 Pruebas de bondad de ajuste sin parámetros especificados

EL CASO DE LOS DATOS DISCRETOS

También podemos realizar una prueba de bondad de ajuste de una hipótesis nula que no especifique completamente las probabilidades $\{p_i, i = 1, \dots, k\}$. Por ejemplo, suponga que estamos interesados en verificar si la cantidad diaria de accidentes de tráfico en cierta región tiene una distribución Poisson con media no especificada. Para verificar la hipótesis, suponga que se obtienen datos durante n días y sea Y_i el número de accidentes en el día i , para $i = 1, \dots, n$. Para determinar si estos datos son congruentes con la hipótesis de una distribución Poisson subyacente, primero debemos enfrentar la dificultad de que, si la hipótesis de Poisson es correcta, estos datos pueden tomar una infinidad de valores posibles. Sin embargo, esto se resuelve dividiendo el conjunto de posibles valores en un número finito, digamos k , de regiones, y luego se decide en cuál de las regiones caen cada uno de los n datos. Por ejemplo, si el área geográfica de interés es pequeña, de modo que haya muy pocos accidentes en un día, podríamos decir que el número de accidentes en cierto día cae en la región i , $i = 1, 2, 3, 4, 5$, si hay $i - 1$

accidentes ese día y en la región 6 si hay cinco o más accidentes. Por lo tanto, si la distribución subyacente es realmente Poisson con media λ , entonces

$$p_i = P\{Y = i - 1\} = \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!}, \quad i = 1, 2, 3, 4, 5 \quad (9.6)$$

$$p_6 = 1 - \sum_{j=0}^4 \frac{e^{-\lambda} \lambda^j}{j!}$$

Otra dificultad que enfrentamos al obtener una prueba de bondad de ajuste de la hipótesis en el sentido de que la distribución subyacente es Poisson, es que el valor medio λ no está dado. Intuitivamente, lo que debemos hacer es estimar su valor a partir de los datos (llamemos $\hat{\lambda}$ a la estimación) y luego calcular el valor del estadístico de prueba

$$T = \sum_{i=1}^k \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

donde N_i es el número de Y_j que caen en la región i , y donde \hat{p}_i es la probabilidad estimada, según H_0 , de que Y_j caiga en la región i , $i = 1, \dots, k$, la cual se obtiene al sustituir $\hat{\lambda}$ en vez de λ en la expresión (9.6).

El método anterior es practicable siempre que se tengan parámetros no especificados en la hipótesis nula, los cuales sean necesarios para calcular las cantidades p_i , $i = 1, \dots, k$. Suponga que hay m parámetros no especificados. Se puede demostrar que, para estimadores razonables de estos parámetros, cuando n es grande la cantidad de prueba T tiene (suponiendo cierta H_0) aproximadamente una distribución ji-cuadrada con $k - 1 - m$ grados de libertad (en otras palabras, se pierde un grado de libertad por cada uno de los parámetros que deben estimarse).

Si la cantidad de prueba toma el valor, digamos, $T = t$, entonces, al servirnos de lo anterior, podemos aproximar el valor p mediante

$$\text{valor } p \approx P\{X_{k-1-m}^2 \geq t\}$$

donde X_{k-1-m}^2 es una variable aleatoria ji-cuadrada con $k - 1 - m$ grados de libertad.

Ejemplo 9d Suponga que en un periodo de 30 días, hubo seis sin accidentes; dos en los cuales ocurrió un accidente; uno en el que sucedieron dos; nueve en los que ocurrieron tres; siete en los que hubo cuatro; cuatro en los que ocurrieron cinco; y uno en el que sucedieron ocho. Para verificar si estos datos son congruentes con la hipótesis de una distribución Poisson subyacente, observe primero que como hubo un total de

87 accidentes, la estimación de la media de la distribución Poisson es

$$\hat{\lambda} = \frac{87}{30} = 2.9$$

Como la estimación de $P\{Y = i\}$ es $e^{-2.9}(2.9)^i/i!$, obtenemos que para las seis regiones dadas al principio de esta sección,

$$\begin{aligned}\hat{p}_1 &= 0.0500, & \hat{p}_2 &= 0.1596, & \hat{p}_3 &= 0.2312, \\ \hat{p}_4 &= 0.2237, & \hat{p}_5 &= 0.1622, & \hat{p}_6 &= 0.1682\end{aligned}$$

Al utilizar los valores $N_1 = 6, N_2 = 2, N_3 = 1, N_4 = 9, N_5 = 7, N_6 = 5$, vemos que el valor del estadístico de prueba es

$$T = \sum_{i=1}^6 \frac{(N_i - 30\hat{p}_i)^2}{30\hat{p}_i} = 19.887$$

Para determinar el valor p , ejecutamos el programa 9-1, que da como resultado

$$\text{valor } p \approx P\{X_4^2 > 19.887\} = 0.0005$$

por lo cual se rechaza la hipótesis de una distribución Poisson subyacente. ■

También podemos utilizar simulación para estimar el valor p . Sin embargo, como la hipótesis nula ya no especifica de manera completa el modelo de probabilidad, el uso de simulación para determinar el valor p del estadístico de prueba es un poco más ingenioso. La forma en que debe hacerse es la siguiente.

- El modelo.* Suponga que la hipótesis nula es que los valores Y_1, \dots, Y_n constituyen una muestra aleatoria de una distribución especificada salvo un conjunto de parámetros desconocidos $\theta_1, \dots, \theta_m$. Suponga también que cuando esta hipótesis es verdadera, los valores posibles de las Y_i son $1, \dots, k$.
- El paso inicial.* Tomamos los datos para estimar los parámetros desconocidos. Específicamente, sea $\hat{\theta}_j$ el valor del estimador de $\theta_j, j = 1, \dots, m$. Ahora, calculamos el valor del estadístico de prueba

$$T = \sum_{i=1}^k \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

donde N_i es el número de valores que son iguales a $i, i = 1, \dots, k$, y \hat{p}_i es la estimación de p_i obtenida al sustituir $\hat{\theta}_j$ en vez de θ_j , para $j = 1, \dots, m$. Sea t el valor de la cantidad de prueba T .

- El paso de simulación.* Ahora realizamos una serie de simulaciones para estimar el valor p de los datos. Primero observe que todas las simulaciones se obtendrán de la distribución poblacional resultante cuando la hipótesis nula es cierta y θ_j es igual a su estimación $\hat{\theta}_j, j = 1, \dots, m$, determinada en el paso (b).

Simulamos una muestra de tamaño n para la distribución poblacional ya mencionada; sea $\hat{\theta}_j$ la estimación de $\theta_j, j = 1, \dots, m$, basada en los datos simulados. A continuación determinamos el valor de

$$T_{\text{sim}} = \sum_{i=1}^k \frac{[N_i - n\hat{p}_i(\text{sim})]^2}{n\hat{p}_i(\text{sim})}$$

donde N_i es el número de valores simulados iguales a $i, i = 1, \dots, k$, y $\hat{p}_i(\text{sim})$ es el valor de p_i cuando θ_j es igual a $\hat{\theta}_j(\text{sim}), j = 1, \dots, m$.

El paso de simulación se debe repetir muchas veces. La estimación del valor p es entonces igual a la proporción de valores de T_{sim} que son al menos tan grandes como t .

Ejemplo 9e Reconsideremos el ejemplo 9d. Los datos presentados produjeron la estimación $\hat{\lambda} = 2.9$ y el valor de la cantidad de prueba $T = 19.887$. Ahora el paso de simulación consiste en generar 30 variables aleatorias independientes de tipo Poisson, cada una con media 2.9 para luego calcular el valor de

$$T^* = \sum_{i=1}^6 \frac{(X_i - 30p_i^*)^2}{30p_i^*}$$

donde X_i es la cantidad de los 30 valores que caen dentro de la región i , y p_i^* es la probabilidad de que una variable aleatoria Poisson con media igual al promedio de los 30 valores generados caiga en la región i . Este paso de simulación se debe repetir muchas veces, y el valor p estimado es la proporción de veces en la que se obtiene un T^* al menos tan grande como 19.887. ■

EL CASO DE LOS DATOS CONTINUOS

Ahora, considere la situación en la que queremos verificar la hipótesis de que las variables aleatorias Y_1, \dots, Y_n tienen la función de distribución continua F_θ , donde $\theta = (\theta_1, \dots, \theta_m)$ es un vector de parámetros desconocidos. Por ejemplo, podríamos estar interesados en verificar que las Y_j provienen de una población con distribución normal. Para emplear la prueba de Kolmogorov-Smirnov, primero empleamos los datos

para estimar el vector de parámetros θ , digamos, mediante el vector de estimadores $\hat{\theta}$. El valor del estadístico de prueba D se calcula mediante

$$D = \underset{x}{\text{Máximo}} |F_e(x) - F_{\hat{\theta}}(x)|$$

donde $F_{\hat{\theta}}$ es la función de distribución obtenida de F_{θ} cuando θ se estima mediante $\hat{\theta}$. Si el valor de la cantidad de prueba es $D = d$, entonces el valor p se puede aproximar de manera burda mediante $P_{F_{\hat{\theta}}}\{D \geq d\} = P_U\{D \geq d\}$. Es decir, después de determinar el valor de D , se obtiene una aproximación burda, que en realidad sobrestima al valor p . Si esto no produce una estimación menor del valor p , entonces, como la hipótesis no sería rechazada, podríamos terminar en este punto. Sin embargo, si este valor p estimado es pequeño, entonces se necesita una forma más precisa de simulación para estimar el verdadero valor p . Describiremos ahora cómo hacerlo.

PASO 1: Utilizamos los datos para estimar θ , digamos, por $\hat{\theta}$. Calculamos el valor de D según lo dicho.

PASO 2: Todas las simulaciones se realizan mediante la distribución $F_{\hat{\theta}}$. Generamos una muestra de tamaño n a partir de esta distribución; sea $\hat{\theta}(\text{sim})$ la estimación de θ con base en esta ejecución de simulación. Calculamos el valor de

$$\underset{x}{\text{Máximo}} |F_{e,\text{sim}}(x) - F_{\hat{\theta}(\text{sim})}(x)|$$

donde $F_{e,\text{sim}}$ es la función de distribución empírica de los datos simulados; observe que es al menos tan grande como d . Repetimos esto muchas veces y utilizamos la proporción de veces que esta cantidad de prueba es al menos tan grande como d como la estimación del valor p .

9.3 El problema de las dos muestras

Suponga que hemos formulado un modelo matemático de un sistema de servicio que elimina todos los clientes al final del día; además, nuestro modelo supone que cada día las leyes de probabilidad son idénticas e independientes. Algunas de las hipótesis individuales del modelo (como por ejemplo que los tiempos de servicio son independientes con una distribución común G o que las llegadas de los clientes constituyen un proceso Poisson) se pueden verificar de manera individual mediante los resultados de las secciones 9.1 y 9.2. Suponga que ninguna de estas pruebas individuales produce un valor p particularmente pequeño, de modo que todas las partes del modelo, consi-

deradas en forma individual, no parecen ser incongruentes con los datos reales del sistema. [Debemos tener cuidado con el significado de un valor p pequeño, pues aunque el modelo fuese correcto, si realizamos un gran número de pruebas tendríamos la probabilidad de que algunos de los valores p resultantes fuesen pequeños. Por ejemplo, si realizamos r pruebas separadas sobre datos independientes, entonces la probabilidad de que al menos uno de los valores p resultantes sea tan pequeño como α es igual a $1 - (1 - \alpha)^r$, lo cual será grande cuando r crezca, aún para α pequeña.]

Sin embargo, en esta etapa no hemos justificado que nuestro modelo sea correcto y válido con base en los datos reales; la totalidad del modelo (incluyendo no sólo las partes individuales, sino las hipótesis acerca de las formas de interacción de estas partes) podría ser impreciso. Una forma de verificar el modelo en su totalidad es considerar cierta cantidad aleatoria que sea una función que tome en cuenta el modelo completo. Por ejemplo, podríamos considerar el tiempo total de espera de todos los clientes que entran al sistema un día determinado. Suponga que hemos observado el sistema real durante m días y sea Y_i , $i = 1, \dots, m$, la suma de estos tiempos de espera para el día i . Si ahora simulamos el modelo matemático propuesto para n días, podemos obtener X_i , $i = 1, \dots, n$ la suma de los tiempos de espera de todos los clientes que llegan el día (simulado) i . Como el modelo matemático supone que todos los días son similares e independientes desde el punto de vista de la probabilidad, se tiene que todas las variables aleatorias X_1, \dots, X_n tienen cierta distribución común, denotada por F . Ahora, si el modelo matemático es una representación precisa del modelo real, entonces los datos reales Y_1, \dots, Y_m también tienen la distribución F . Es decir, si el modelo matemático es preciso, uno no podría distinguir los datos simulados de los datos reales. Esto implica que una forma de verificar la precisión del modelo en su totalidad es verificar la hipótesis nula H_0 de que $X_1, \dots, X_n, Y_1, \dots, Y_m$ son variables aleatorias independientes con una distribución común. Ahora mostraremos la forma de verificar esta hipótesis.

Suponga que tenemos dos conjuntos de datos X_1, \dots, X_n y Y_1, \dots, Y_m y que queremos verificar la hipótesis H_0 de que estas $n + m$ variables aleatorias son todas independientes e idénticamente distribuidas. Este problema de verificación de una hipótesis estadística se conoce como el problema de dos muestras.

Para verificar H_0 , ordenamos los $n + m$ valores $X_1, \dots, X_n, Y_1, \dots, Y_m$ y suponemos por el momento que todos los valores $n + m$ son distintos, de modo que el orden es único. Para $i = 1, \dots, n$, sea R_i el rango de X_i entre los $n + m$ valores; es decir, $R_i = j$ si X_i es el j -ésimo valor más pequeño entre los $n + m$ valores. La cantidad

$$R = \sum_{i=1}^n R_i$$

igual a la suma de los rangos del primer conjunto de datos, se utiliza como cantidad de prueba (cualquiera de los dos conjuntos de datos se puede considerar como el "primero").

Si R es demasiado grande (lo cual indica que el primer conjunto de datos tiende a ser mayor que el segundo) o muy pequeño (que indica lo contrario), entonces esto sería una fuerte evidencia en contra de la hipótesis nula. Específicamente, si $R = r$, rechazamos la hipótesis nula si

$$P_{H_0}\{R \leq r\} \quad \text{o} \quad P_{H_0}\{R \geq r\}$$

es demasiado baja. De hecho, el valor p de los datos de prueba que producen $R = r$ está dado por

$$\text{valor } p = 2 \text{ Mínimo } (P_{H_0}\{R \leq r\}, P_{H_0}\{R \geq r\}) \quad (9.7)$$

[Elegimos el doble del mínimo de las probabilidades pues rechazamos la hipótesis si R es demasiado pequeña o demasiado grande. Por ejemplo, supongamos que r_* y r^* son tales que la probabilidad bajo H_0 de obtener un valor menor (mayor) o igual a r_* (r^*) es 0.05. Como la probabilidad de que ocurra cualquiera de estos eventos bajo H_0 es 0.1, tenemos que si el resultado es r_* (o r^*), el valor p es 0.1.]

La prueba de hipótesis resultante del valor p anterior (es decir, la prueba que rechaza la hipótesis nula cuando el valor p es suficientemente pequeño) es la *prueba de la suma de rangos de dos muestras* (también llamada prueba de dos muestras de Wilcoxon y prueba de dos muestras de Mann-Whitney).

Ejemplo 9f Suponga que una observación directa de un sistema durante cinco días muestra que cierta cantidad toma los valores sucesivos

342, 448, 504, 361, 453

mientras que una simulación de 10 días de un modelo matemático propuesto para el sistema produjo los siguientes valores:

186, 220, 225, 456, 276, 199, 371, 426, 242, 311

Como los cinco valores del primer conjunto tienen rangos 8, 12, 15, 9, 13, tenemos que el valor de la cantidad de prueba es $R = 57$. ■

Podemos calcular en forma explícita el valor p dado en la ecuación (9.7) cuando n y m no son demasiado grandes y todos los datos son distintos. Para esto, sea

$$P_{n,m}(r) = P_{H_0}\{R \leq r\}$$

Por lo tanto, $P_{n,m}(r)$ es la probabilidad de que de dos conjuntos de datos idénticamente distribuidos de tamaños n y m , la suma de los rangos de los valores del primer conjunto

sea menor o igual a r . Podemos obtener una ecuación recursiva para estas probabilidades condicionando si el máximo valor proviene del primero o del segundo conjunto. Si el valor máximo está dentro del primer conjunto de datos, la suma de sus rangos es igual a $n + m$ (el rango del valor máximo) más la suma de los rangos de los otros $n - 1$ valores del conjunto, considerados junto con los m valores del otro conjunto. Por lo tanto, cuando el máximo está dentro del primer conjunto de datos, la suma de los rangos de ese conjunto es menor o igual a r si la suma de los rangos de los $n - 1$ elementos restantes es menor o igual a $r - n - m$, y esto es cierto con probabilidad $P_{n-1,m}(r - n - m)$. Con un argumento similar se muestra que si el valor máximo está dentro del segundo conjunto, la suma de los rangos del primer conjunto es menor o igual a r con probabilidad $P_{n,m-1}(r)$. Por último, como el valor máximo tiene la misma probabilidad de ser cualquiera de los $n + m$ valores, será un miembro del primer conjunto con probabilidad $n/(n + m)$. Al reunir todo lo anterior obtenemos la siguiente ecuación recursiva:

$$P_{n,m}(r) = \frac{n}{n + m} P_{n-1,m}(r - n - m) + \frac{m}{n + m} P_{n,m-1}(r) \quad (9.8)$$

Si partimos de las condiciones de frontera

$$P_{1,0}(k) = \begin{cases} 0, & k \leq 0 \\ 1, & k > 0 \end{cases} \quad \text{y} \quad P_{0,1}(k) = \begin{cases} 0, & k < 0 \\ 1, & k \geq 0 \end{cases}$$

resolvemos la ecuación (9.8) en forma recursiva para obtener $P_{n,m}(r) = P_{H_0}\{R \leq r\}$ y $P_{n,m}(r - 1) = 1 - P_{H_0}\{R \geq r\}$.

El programa 9-4 utiliza la recursión de la ecuación (9.8) para calcular el valor p para la prueba de suma de rangos. Las entradas necesarias son los tamaños del primer y segundo conjuntos de datos y la suma de los rangos de los elementos del primer conjunto. Aunque cualquier conjunto se puede designar como el primero, el programa es más rápido si el primer conjunto es aquel cuya suma de rangos es menor.

Ejemplo 9g Cinco días de observación de un sistema produjeron los siguientes valores de cierta cantidad de interés:

132, 104, 162, 171, 129

Una simulación de 10 días de un modelo propuesto para este sistema arrojó los valores

107, 94, 136, 99, 114, 122, 108, 130, 106, 88

Suponga que el modelo formulado implica que estos valores diarios deben ser independientes y tener una distribución común. Para determinar el valor p resultante de

los datos anteriores, observe primero que R , la suma de los rangos de la primera muestra, es

$$R = 12 + 4 + 14 + 15 + 10 = 55$$

Para determinar el valor p utilizamos el programa 9-4, el cual produce lo siguiente:

```

RUN
ESTE PROGRAMA CALCULA EL VALOR p PARA LA PRUEBA DE SUMA DE
RANGOS DE DOS MUESTRAS
ESTE PROGRAMA SERÁ MÁS RÁPIDO SI USTED DESIGNA COMO PRIMERA
MUESTRA AQUELLA CON MENOR SUMA DE RANGOS
PROPORCIONE EL TAMAÑO DE LA PRIMERA MUESTRA
? 5
PROPORCIONE EL TAMAÑO DE LA SEGUNDA MUESTRA
? 10
PROPORCIONE LA SUMA DE LOS RANGOS DE LA PRIMERA MUESTRA
? 55
EL VALOR p ES 0.0752579
OK

```

La dificultad en el uso de la recursión (9.8) para calcular el valor p es que la cantidad de cálculos crece de manera enorme cuando la muestra aumenta de tamaño. Por ejemplo, si $n = m = 20$, aunque elijamos la cantidad de prueba como la menor suma de rangos, entonces, como la suma de todos los rangos es $1 + 2 + \dots + 40 = 820$, es posible que el estadístico de prueba tenga un valor tan grande como 410. Por lo tanto, puede haber hasta $20 \times 20 \times 410 = 164,000$ valores de $P_{n,m}(r)$ que habría que calcular para determinar el valor p . Así, para muestras de gran tamaño, el uso de la recursión proporcionada por (9.8) no sería viable. Otros métodos de aproximación que pueden utilizarse en estos casos son (a) un método clásico basado en una aproximación de la distribución de R y (b) la simulación.

En el método clásico de aproximación del valor p , basados en el hecho de que según H_0 todos los posibles ordenamientos de los $n + m$ valores son igualmente probables. Este hecho permite mostrar fácilmente que

$$E_{H_0}[R] = n \frac{(n + m + 1)}{2}$$

$$\text{Var}_{H_0}(R) = nm \frac{(n + m + 1)}{12}$$

Se puede mostrar que según H_0 , cuando n y m son grandes, R se distribuye normal, aproximadamente. Por lo tanto, cuando H_0 es cierta,

$$\frac{R - n(n + m + 1)/2}{\sqrt{nm(n + m + 1)/12}} \text{ es aproximadamente una normal estándar.}$$

Como para una variable aleatoria normal W , el mínimo de $P\{W \leq r\}$ y $P\{W \geq r\}$ es la primera cuando $r \leq E[W]$, y la segunda en caso contrario, tenemos que cuando n y m no son demasiado pequeñas (basta que ambas sean mayores que 7), podemos aproximar el valor p del resultado de la prueba $R = r$ mediante

$$\text{valor } p \approx \begin{cases} 2 P\{Z < r^*\} & \text{si } r \leq n \frac{(n + m + 1)}{2} \\ 2 P\{Z > r^*\} & \text{en caso contrario} \end{cases} \quad (9.9)$$

donde

$$r^* = \frac{r - \frac{n(n + m + 1)}{2}}{\sqrt{\frac{nm(n + m + 1)}{12}}}$$

y Z es una variable aleatoria normal estándar.

Ejemplo 9h Veamos cómo funciona la aproximación clásica para los datos del ejemplo 9g. En este caso, como $n = 5$ y $m = 10$, tenemos que

$$\text{valor } p = 2 P_{H_0}\{R \geq 55\}$$

$$\begin{aligned} &\approx 2 P\left\{Z \geq \frac{55 - 40}{\sqrt{\frac{50 \times 16}{12}}}\right\} \\ &= 2 P\{Z \geq 1.8371\} \\ &= 0.066 \end{aligned}$$

que debemos comparar con la respuesta exacta 0.075. ■

El valor p de la prueba de rango de dos muestras también se puede aproximar mediante una simulación. Para esto, recuerde que si el valor observado de la cantidad de prueba R es $R = r$, entonces el valor p está dado por

$$\text{valor } p = 2 \text{ Mínimo } (P_{H_0}\{R \geq r\}, P_{H_0}\{R \leq r\})$$

Ahora, según H_0 , siempre que los $n + m$ valores sean distintos, se tiene que todos los ordenamientos entre estos valores son igualmente probables, y así los rangos del primer conjunto de datos de tamaño n tienen la misma distribución de una selección aleatoria de n de los valores $1, 2, \dots, n + m$. Así, según H_0 , para aproximar la distribución de probabilidad de R se simula en forma continua un subconjunto aleatorio de n de los enteros $1, 2, \dots, n + m$ y se determina la suma de los elementos en el subconjunto. El valor de $P_{H_0}\{R \leq r\}$ se puede aproximar mediante la proporción de simulaciones que produzcan una suma menor o igual a r , y el valor de $P_{H_0}\{R \geq r\}$ mediante la proporción de simulaciones que produzcan una suma mayor o igual a r .

El programa 9-5 aproxima el valor p realizando la simulación anterior. Opera sumando en forma continua los últimos n términos de una permutación aleatoria de los números $1, 2, \dots, n + m$ (el método utilizado para simular una permutación aleatoria se explica en el ejemplo 3a del capítulo 3). El programa es más eficiente si el conjunto de datos de menor tamaño se designa como primera muestra.

Ejemplo 9i Al utilizar el programa 9-5 con 1000 ejecuciones para los datos del ejemplo 9g se obtiene lo siguiente:

```

RUN
ESTE PROGRAMA APROXIMA EL valor p EN LA PRUEBA DE SUMA DE
RANGOS DE DOS MUESTRAS MEDIANTE UN ESTUDIO DE SIMULACIÓN
Random number seed (-32768 a 32767)? 4566
PROPORCIONE EL TAMAÑO DE LA PRIMERA MUESTRA
? 5
PROPORCIONE EL TAMAÑO DE LA SEGUNDA MUESTRA
? 10
PROPORCIONE LA SUMA DE LOS RANGOS DE LA PRIMERA MUESTRA
? 55
PROPORCIONE EL NÚMERO DESEADO DE EJECUCIONES DE LA SIMU-
LACIÓN
? 1000
EL valor p APROXIMADO ES 0.07
OK

```

El análisis anterior supone que todos los $n + m$ valores son distintos. Cuando algunos de los valores tienen un valor común, uno debe considerar como el rango de un valor el promedio de los rangos de los valores iguales a él. Por ejemplo, si el primer conjunto de datos es 2, 3, 4 y el segundo es 3, 5, 7, entonces la suma de los rangos del primer conjunto es $1 + 2.5 + 4 = 7.5$. El valor p se debe aproximar mediante la aproximación normal, mediante la ecuación (9.9).

Una generalización del problema de dos muestras es el problema de varias muestras, en el que se tienen los siguientes m conjuntos de datos:

$$\begin{array}{cccc} X_{1,1}, & X_{1,2}, & \dots, & X_{1,n_1} \\ X_{2,1}, & X_{2,2}, & \dots, & X_{2,n_2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m,1}, & X_{m,2}, & \dots, & X_{m,n_m} \end{array}$$

y estamos interesados en verificar la hipótesis nula H_0 de que todas las $n = \sum_{i=1}^m n_i$ variables aleatorias son independientes y tienen una distribución común. Una generalización de la prueba de rango de dos muestras, llamada la prueba de rango de varias muestras (o bien prueba de Kruskal-Wallis), se obtiene estableciendo primero el rango de los n valores. Luego, sea $R_i, i = 1, \dots, m$ la suma de los rangos de los n_i valores datos del i -ésimo conjunto (observe que con esta notación, R_i es una suma de rangos y no un rango único como antes). Como según H_0 todos los ordenamientos son igualmente probables (siempre que los valores sean distintos), tenemos exactamente igual que antes que

$$E[R_i] = n_i \frac{(n+1)}{2}$$

Dado lo anterior, la prueba de suma de rangos de varias muestras se basa en la cantidad de prueba

$$R = \frac{12}{n(n+1)} \sum_{i=1}^m \frac{[R_i - n_i(n+1)/2]^2}{n_i}$$

Como los valores pequeños de R indican un buen ajuste a H_0 , la prueba basada en la cantidad R rechaza a H_0 para valores suficientemente grandes de R . De hecho, si el valor observado de R es $R = y$, el valor p de este resultado está dado por

$$\text{valor } p = P_{H_0}\{R \geq y\}$$

Este valor se puede aproximar utilizando el resultado de que para valores grandes de n_1, \dots, n_m , R tiene aproximadamente una distribución ji-cuadrada con $m - 1$ grados de libertad [este último resultado es la razón de la inclusión del término $12/n(n+1)$ en la definición de R]. Por lo tanto, si $R = y$,

$$\text{valor } p \approx P\{X_{m-1}^2 \geq y\}$$

La simulación también sirve para evaluar el valor p (véase el ejercicio 14).

Aunque los valores no sean todos distintos, hay que utilizar la aproximación anterior del valor p . Al calcular el valor de R , el rango de un valor individual debe ser, como antes, el promedio de todos los rangos de los datos iguales a él.

9.4 Validación de la hipótesis de un proceso Poisson no homogéneo

Considere un modelo matemático que supone que las llegadas diarias a un sistema ocurren de acuerdo con un proceso Poisson no homogéneo, independiente de un día a otro y que tiene una función de intensidad común no especificada.

Para establecer la validez de esta hipótesis, suponga que observamos el sistema durante r días, anotando los tiempos de llegada. Sea N_i , $i = 1, \dots, r$ el número de llegadas el día i y observe que si el proceso de llegada es un proceso Poisson no homogéneo, entonces estas cantidades son variables aleatorias Poisson, independientes y con la misma media. Aunque esta consecuencia se podría verificar mediante el método de bondad de ajuste, como en el ejemplo 9a, presentamos otro método que a veces es más eficaz. Este método se basa en el hecho de que la media y la varianza de una variable aleatoria Poisson son iguales. Por lo tanto, si los N_i surgen realmente de una distribución Poisson, la media muestral

$$\bar{N} = \sum_{i=1}^r \frac{N_i}{r}$$

y la varianza muestral

$$S^2 = \sum_{i=1}^r \frac{(N_i - \bar{N})^2}{r-1}$$

deben ser aproximadamente iguales. A partir de esto, basamos nuestra prueba de la hipótesis H_0 : N_i son variables aleatorias Poisson independientes con la misma media, en la cantidad de prueba

$$T = \frac{S^2}{\bar{N}} \quad (9.10)$$

Ya que un valor muy pequeño o muy grande T sería incongruente con H_0 , el valor p para el resultado $T = t$ sería

$$\text{valor } p = 2 \text{ Mínimo } P_{H_0}\{T \leq t\}, P_{H_0}\{T \geq t\}$$

Sin embargo, como H_0 no especifica la media de la distribución Poisson, no podemos calcular de manera inmediata las probabilidades anteriores, sino que primero

debemos utilizar los datos observados para estimar la media. Utilizando el estimador \bar{N} , se concluye que si el valor observado de \bar{N} es $\bar{N} = m$, entonces el valor p se puede aproximar mediante

$$\text{valor } p \approx 2 \text{ Mínimo } (P_m\{T \leq t\}, P_m\{T \geq t\})$$

donde T se define en la ecuación (9.10), con N_1, \dots, N_r variables aleatorias Poisson independientes, cada una con media m . Ahora podemos aproximar $P_m\{T \leq t\}$ y $P_m\{T \geq t\}$ mediante una simulación. Es decir, generamos de manera continua r variables aleatorias Poisson independientes con media m y calculamos el valor resultante de T . La proporción de estas variables para las cuales $T \leq t$ es nuestra estimación de $P\{T \leq t\}$ y la proporción de estas variables para las cuales $T \geq t$ es nuestra estimación de $P\{T \geq t\}$.

Si el valor p anterior es muy pequeño, rechazamos la hipótesis nula de que las llegadas diarias constituyen un proceso Poisson no homogéneo. Sin embargo, si el valor p no es pequeño, esto sólo implica que la hipótesis de que el número de llegadas cada día tenga una distribución Poisson es una hipótesis viable, y en sí mismo no establece la validez de la hipótesis más fuerte en el sentido de que el patrón real de llegadas (determinado por la función de intensidad no homogénea) sea el mismo de un día a otro. Para completar nuestra validación, debemos considerar los tiempos reales de llegada para cada uno de los r días observados. Suponga que los tiempos de llegada el día j , $j = 1, \dots, r$, son $X_{j,1}, X_{j,2}, \dots, X_{j,N_j}$. Ahora, si el proceso de llegada es realmente un proceso Poisson no homogéneo, se puede mostrar que cada uno de estos r conjuntos de tiempos de llegada constituye una muestra de una distribución común. Es decir, según la hipótesis nula, los r conjuntos de datos $X_{j,1}, \dots, X_{j,N_j}$, $j = 1, \dots, r$ son variables aleatorias independientes con una distribución común.

Sin embargo, la consecuencia anterior se puede verificar mediante la prueba de rango de varias muestras dada en la sección 9.3. Es decir, primero establecemos el rango de los $N \equiv \sum_{j=1}^r N_j$ valores; sea R_j la suma de los rangos de los N_j valores del j -ésimo conjunto. La cantidad de prueba

$$R = \frac{12}{N(N+1)} \sum_{j=1}^r \frac{\left(R_j - N_j \frac{(N+1)}{2}\right)^2}{N_j}$$

se puede emplear utilizando el hecho de que, cuando H_0 es cierta, R tiene aproximadamente una distribución ji-cuadrada con $r-1$ grados de libertad. Por lo tanto, si el valor observado de R es $R = y$, el valor p resultante se puede aproximar mediante

$$\begin{aligned} \text{valor } p &= 2 \text{ mínimo } (P_{H_0}\{R \leq y\}, P_{H_0}\{R \geq y\}) \\ &\approx 2 \text{ mínimo } (P\{X_{r-1}^2 \leq y\}, 1 - P\{X_{r-1}^2 \leq y\}) \end{aligned}$$

donde X_{r-1}^2 es una variable aleatoria ji-cuadrada con $r - 1$ grados de libertad (por supuesto, también podríamos aproximar el valor p mediante una simulación). Si este valor p , junto con el valor p antes analizado, no es demasiado pequeño, podemos concluir que los datos no son incongruentes con nuestra hipótesis de que las llegadas diarias constituyen un proceso Poisson no homogéneo.

Ejemplo 9j Suponga que los tiempos diarios de entrega en cierta planta se anotan durante cinco días. Durante este tiempo, los números de entregas en cada uno de los días son los siguientes:

18, 24, 16, 19, 25

Suponga además que cuando los 102 tiempos de entrega se ordenan de acuerdo con la hora del día en que llegaron, las sumas de los rangos de las entregas de cada día son

1010, 960, 1180, 985, 1118

Con los datos anteriores, verificaremos la hipótesis de que el proceso de llegada diaria de las entregas es un proceso Poisson no homogéneo.

Primero verificamos que el primer conjunto de datos del número diario de entregas conste de un conjunto de cinco variables aleatorias Poisson independientes e idénticamente distribuidas. La media y la varianza muestrales son iguales a

$$\bar{N} = 20.4 \quad y \quad S^2 = 15.3$$

de modo que el valor de la cantidad de prueba es $T = 0.75$. Para determinar el valor p aproximado de la prueba de que las N_i son variables aleatorias Poisson independientes, simulamos 500 conjuntos de cinco variables aleatorias Poisson con media 20.4 y luego calculamos el valor resultante de $T = S^2/\bar{N}$. La salida de esta simulación indica un valor p aproximado de 0.84, de modo que es claro que la hipótesis de que los números de entregas diarias son variables aleatorias Poisson independientes con una media común es congruente con los datos.

Para continuar nuestra prueba de la hipótesis nula de un proceso Poisson no homogéneo, calculamos el valor de la cantidad de prueba R , que es igual a 14.425. Como la probabilidad de que una variable aleatoria ji-cuadrada con cuatro grados de libertad sea tan grande como 14.425 es (del programa 9-1) 0.006, tenemos que el valor p es 0.012. Para un valor p tan pequeño, debemos rechazar la hipótesis nula. ■

Si hubiéramos querido verificar la hipótesis de que el proceso de llegadas diarias constituía un proceso Poisson homogéneo, procederíamos como antes y primero verificaríamos la hipótesis de que el número de llegadas cada día son variables aleatorias Poisson independientes e idénticamente distribuidas. Si la hipótesis sigue siendo plau-

sible después de realizar esta prueba, de nuevo continuamos como en el caso no homogéneo, considerando el conjunto real de $N = \sum_{j=1}^r N_j$ tiempos de llegada. Sin embargo, ahora utilizamos el resultado de que, en un proceso Poisson homogéneo, se puede mostrar que, dado el número de llegadas en un día, las llegadas reales son independientes y están distribuidas de manera uniforme en $(0, T)$, donde T es la longitud de un día. Sin embargo, esta consecuencia se puede verificar mediante la prueba de bondad de ajuste de Kolmogorov-Smirnov presentada en la sección 9.1. Es decir, si las llegadas constituyen un proceso Poisson homogéneo, podemos considerar que las N variables aleatorias $X_{j,i}$, $i = 1, \dots, N_j$, $j = 1, \dots, r$, donde $X_{j,i}$ representa el i -ésimo tiempo de llegada del día j , forman un conjunto de N variables aleatorias independientes, distribuidas de manera uniforme en $(0, T)$. Por lo tanto, si definimos la función de distribución empírica F_e haciendo que $F_e(x)$ sea la proporción de los N valores que son menores o iguales a x ; es decir,

$$F_e(x) = \sum_{j=1}^r \sum_{i=1}^{N_j} \frac{I_{j,i}}{N}$$

donde

$$I_{j,i} = \begin{cases} 1 & \text{si } X_{j,i} \leq x \\ 0 & \text{en caso contrario} \end{cases}$$

entonces el valor de la cantidad de prueba es

$$D = \begin{matrix} \text{Máximo} \\ 0 \leq x \leq T \end{matrix} \left| F_e(x) - \frac{x}{T} \right|$$

Una vez determinado el valor del estadístico de prueba D , podemos determinar el valor p resultante mediante una simulación, como en la sección 9.1.

Si se muestra que la hipótesis de un proceso Poisson no homogéneo es congruente con los datos, ahora enfrentamos el problema de estimar la función de intensidad $\lambda(t)$, $0 \leq t \leq T$, de este proceso [en el caso homogéneo, el estimador obvio es $\lambda(t) = \hat{\lambda}/T$ cuando $\hat{\lambda}$ es la estimación del número medio de llegadas en un día de longitud T]. Para estimar la función de intensidad, ordenamos los $N = \sum_{j=1}^r N_j$ tiempos de llegada de cada día. Sea $y_0 = 0$, y para $k = 1, \dots, N$, sea y_k el k -ésimo valor más pequeño de estos N tiempos de llegada. Como existe un total de una llegada en r días dentro del intervalo de tiempo (y_{k-1}, y_k) , $k = 1, \dots, N$, una estimación razonable de $\lambda(t)$ sería

$$\hat{\lambda}(t) = \frac{1}{r(y_k - y_{k-1})} \quad \text{para } y_{k-1} < t < y_k$$

[Para entender el estimador anterior, observe que si $\hat{\lambda}(t)$ fuese la función de intensidad, el número esperado de llegadas diarias que ocurrirían en un instante t tal que $y_{k-1} < t \leq y_k$ estaría dado por

$$E[N(y_k) - N(y_{k-1})] = \int_{y_{k-1}}^{y_k} \hat{\lambda}(t) dt = \frac{1}{r}$$

y por tanto, el número esperado de llegadas dentro de ese intervalo durante r días sería 1, lo cual coincide con el número real de llegadas observado en ese intervalo.]

Ejercicios

- De acuerdo con la teoría genética de Mendel, cierta planta de guisantes debe producir flores blancas, rosas o rojas con probabilidades respectivas $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$. Para verificar la teoría, se estudió una muestra de 564 guisantes, con el resultado de que 141 produjeron flores blancas, 291 flores rosas y 132 flores rojas. Aproxime el valor p de este conjunto de datos
 - utilizando la aproximación ji-cuadrada, y
 - realizando una simulación.
- Para verificar si cierto dado no estaba trucado, se registraron 1000 lanzamientos del dado, con el resultado de que el número de veces que el dado cayó en i , $i = 1, 2, 3, 4, 5, 6$ fueron, respectivamente, 158, 172, 164, 181, 160, 165. Aproxime el valor p de la prueba de que el dado no está trucado
 - utilizando la aproximación ji-cuadrada, y
 - utilizando simulación.
- Aproxime el valor p de la hipótesis de que los siguientes 10 valores son números aleatorios: 0.12, 0.18, 0.06, 0.33, 0.72, 0.83, 0.36, 0.27, 0.77, 0.74.
- Aproxime el valor p de la hipótesis de que el siguiente conjunto de datos de 14 puntos es una muestra de una distribución uniforme en (50, 200):

164, 142, 110, 153, 103, 52, 174, 88, 178, 184, 58, 62, 132, 128
- Aproxime el valor p de la hipótesis de que los siguientes 13 datos provienen de una distribución exponencial con media 50:

86, 133, 75, 22, 11, 144, 78, 122, 8, 146, 33, 41, 99

- Aproxime el valor p de la prueba de que los siguientes datos provienen de una distribución binomial con parámetros $(8, p)$, donde p no se conoce:

6, 7, 3, 4, 7, 3, 7, 2, 6, 3, 7, 8, 2, 1, 3, 5, 8, 7
- Aproxime el valor p de la prueba de que el siguiente conjunto de datos proviene de una población con distribución exponencial: 122, 133, 106, 128, 135, 126.
- Para generar los valores ordenados de n números aleatorios podríamos generar n números aleatorios y luego ordenarlos. Otro método utiliza el resultado de que dado que el $(n+1)$ -ésimo evento de un proceso Poisson ocurre en el instante t , los primeros n tiempos de evento se distribuyen como el conjunto de valores ordenados de n variables aleatorias uniformes en $(0, t)$. Use este resultado para explicar por qué, en el siguiente algoritmo, y_1, \dots, y_n denotan los valores ordenados de n números aleatorios.

Generar $n+1$ números aleatorios U_1, \dots, U_{n+1}

$$X_i = -\log U_i, \quad i = 1, \dots, n+1$$

$$t = \sum_{i=1}^{n+1} X_i, \quad c = \frac{1}{t}$$

$$y_i = y_{i-1} + cX_i, \quad i = 1, \dots, n \text{ (con } y_0 = 0)$$

- Genere los valores de 10 variables aleatorias exponenciales independientes, cada una con media 1. Luego, con base en la cantidad de prueba de Kolmogorov-Smirnov, aproxime el valor p de la prueba de que los datos realmente provienen de una distribución exponencial con media 1.
- Un experimento diseñado para comparar dos tratamientos contra la corrosión arrojó los siguientes datos (los cuales representan la máxima profundidad de los agujeros en unidades de milésimas de pulgada) en pedazos de alambre sujetos a uno de los dos tratamientos:

Tratamiento 1:	65.2	67.1	69.4	78.4	74.0	80.3
Tratamiento 2:	59.4	72.1	68.0	66.2	58.5	

Calcule el valor p exacto de este conjunto de datos, al verificar la hipótesis de que ambos tratamientos tienen resultados idénticos.
- En el ejercicio 10, calcule el valor p aproximado con base en
 - la aproximación normal, y
 - una simulación.
- Catorce ciudades, aproximadamente del mismo tamaño, se eligen para un estudio de seguridad vial. Si siete de ellas se eligen al azar y durante un mes aparecen en los periódicos locales artículos relativos a la seguridad vial. Los

números de accidentes de tránsito del mes posterior a la campaña son los siguientes:

Grupo de tratamiento:	19	31	39	45	47	66	75
Grupo de control:	28	36	44	49	52	72	72

Determine el valor p exacto al verificar la hipótesis de que los artículos no tuvieron efecto alguno.

13. Aproxime el valor p del ejercicio 12
 - (a) utilizando la aproximación normal, y
 - (b) utilizando una simulación.
14. Explique cómo se puede utilizar la simulación para aproximar el valor p en el problema de varias muestras; es decir, cuando se verifica que un conjunto de m muestras provienen todas de la misma distribución de probabilidad.
15. Considere los siguientes datos resultantes de tres muestras:

Muestra 1:	121	144	158	169	194	211	242
Muestra 2:	99	128	165	193	242	265	302
Muestra 3:	129	134	137	143	152	159	170

Calcule el valor p aproximado de la prueba de que todos los datos provienen de una única distribución de probabilidad

 - (a) utilizando la aproximación ji-cuadrada, y
 - (b) utilizando una simulación.
16. El número de llegadas diarias durante un periodo de ocho días es el siguiente:

122, 118, 120, 116, 125, 119, 124, 130

¿Cree usted que las llegadas diarias pueden ser independientes e idénticamente distribuidas como procesos Poisson no homogéneos?
17. Durante un intervalo de longitud 100, han ocurrido 18 llegadas en los siguientes tiempos:

12, 20, 33, 44, 55, 56, 61, 63, 66, 70, 73, 75, 78, 80, 82, 85, 87, 90

Aproxime el valor p de la prueba de que el proceso de llegada es un proceso Poisson (homogéneo).

Bibliografía

Diaconis, P. y B. Efron, "Computer Intensive Methods in Statistics", en *Scientific American*, **248**:(5), 1983, pp.96-109.

- Fishman, G. S., *Concepts and Methods in Discrete Event Digital Simulations*, Nueva York, Wiley, 1973.
- Kendall, M. y A. Stuart, *The Advanced Theory of Statistics*, Nueva York, MacMillan, 4a. ed., 1979.
- Mihram, G. A., *Simulation — Statistical Foundations and Methodology*, Nueva York, Academic Press, 1972.
- Sargent, R. G., "A Tutorial on Validation and Verification of Simulation Models", en *Proc. 1988 Winter Simulation Conf.*, San Diego, 1988, pp. 33-39.
- Schruben, L. W., "Establishing the Credibility of Simulations", en *Simulation*, **34**, 1980, 101-105.

Métodos de Monte Carlo con cadenas de Markov

Introducción

En general, es difícil simular el valor de un vector aleatorio \mathbf{X} cuyas variables aleatorias componentes son dependientes. En este capítulo presentamos un método poderoso para generar un vector cuya distribución es aproximadamente la de \mathbf{X} . Este método, llamado método de Monte Carlo con cadenas de Markov, tiene la ventaja adicional de que la función de masa (o densidad) de \mathbf{X} puede estar dada salvo una constante multiplicativa, lo cual es de gran importancia en las aplicaciones, como veremos más adelante.

En la sección 10.1 presentamos los resultados necesarios relativos a las cadenas de Markov. En la sección 10.2 explicamos el algoritmo de Hastings-Metropolis para construir una cadena de Markov con una función de masa de probabilidad dada como distribución límite. En la sección 10.3, vemos un caso particular de este algoritmo, el muestreador de Gibbs, que es tal vez el método de Monte Carlo con cadenas de Markov más utilizado. Estudiamos en la sección 10.4 una aplicación de los métodos anteriores a los problemas de optimización determinista, conocida como temple simulado. En la sección 10.5 presentamos la técnica de muestreo con remuestreo de importancia (*sampling importance resampling*, SIR). Aunque no es estrictamente un algoritmo de Monte Carlo con cadenas de Markov, también permite simular en forma aproximada un vector aleatorio cuya función de masa está dada salvo una constante multiplicativa.

10.1 Cadenas de Markov

Consideremos una colección de variables aleatorias X_0, X_1, \dots . Interpretamos X_n como el "estado del sistema en el instante n " y suponga que el conjunto de valores posibles de las

X_n (es decir, los posibles estados del sistema) son el conjunto $1, \dots, N$. Si hay un conjunto de números P_{ij} , $i, j = 1, \dots, N$, tales que siempre que el proceso esté en el estado i , entonces, de manera independiente a los estados pasados, la probabilidad de que el siguiente estado sea j es P_{ij} , entonces decimos que la colección $\{X_n, n \geq 0\}$ es una *cadena de Markov* con probabilidades de transición P_{ij} , $i, j = 1, \dots, N$. Como el proceso debe estar en algún estado después de dejar el estado i , estas probabilidades de transición satisfacen

$$\sum_{j=1}^N P_{ij} = 1, \quad i = 1, \dots, N$$

Una cadena de Markov es irreducible si para cada par de estados i y j existe una probabilidad positiva de que, partiendo del estado i , el proceso llega a estar en el estado j . Para una cadena de Markov irreducible, sea π_j la fracción de tiempo, a largo plazo, que el proceso pasa en el estado j (se puede mostrar que π_j existe y es constante, con probabilidad 1, de manera independiente al estado inicial). Es posible mostrar que las cantidades π_j , $j = 1, \dots, N$, son la única solución del siguiente conjunto de ecuaciones lineales:

$$\pi_j = \sum_{i=1}^N \pi_i P_{ij}, \quad j = 1, \dots, N \quad (10.1)$$

$$\sum_{j=1}^N \pi_j = 1$$

Observación El conjunto de ecuaciones (10.1) tiene una interpretación heurística. Como π_i es la fracción de tiempo que la cadena de Markov está en el estado i y como cada transición que sale del estado i está en el estado j con probabilidad P_{ij} , se tiene que $\pi_i P_{ij}$ es la fracción de tiempo en la cual la cadena de Markov acaba de entrar al estado j desde el estado i . Por lo tanto, la primera ecuación de (10.1) establece el hecho intuitivamente claro de que la fracción de tiempo en la cual la cadena de Markov acaba de entrar al estado j es igual a la suma, sobre todos los estados i , de la fracción de tiempo en que acaba de entrar al estado j proveniente del estado i . Por supuesto, la segunda ecuación de (10.1) dice que la suma de la fracción de tiempo en la cual la cadena está en el estado j , sobre toda j , debe ser igual a 1. ■

Con frecuencia las $\{\pi_j\}$ se llaman *probabilidades estacionarias* de la cadena de Markov, ya que si el estado inicial de la cadena de Markov se distribuye de acuerdo con $\{\pi_j\}$, entonces $P\{X_n = j\} = \pi_j$, para toda n y j (véase el ejercicio 1).

Una propiedad importante de las cadenas de Markov es que para cualquier función h sobre el espacio de estados, con probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N \pi_j h(j) \quad (10.2)$$

Lo anterior se debe a que si $p_j(n)$ es la fracción de tiempo que la cadena está en el estado j entre los instantes $1, \dots, n$, entonces

$$\frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N h(j)p_j(n) \rightarrow \sum_{j=1}^N h(j)\pi_j$$

Con frecuencia, la cantidad π_j se puede interpretar como la probabilidad límite de que la cadena esté en el estado j . Para precisar las condiciones en las que es válida esta interpretación, primero necesitamos la definición de una cadena de Markov aperiódica.

Definición Una cadena de Markov irreducible es aperiódica si para algún $n \geq 0$ y algún estado j ,

$$P\{X_n = j | X_0 = j\} > 0 \quad \text{y} \quad P\{X_{n+1} = j | X_0 = j\} > 0$$

Se puede mostrar que si la cadena de Markov es irreducible y aperiódica, entonces

$$\pi_j = \lim_{n \rightarrow \infty} P\{X_n = j\}, \quad j = 1, \dots, N$$

A veces hay una forma más fácil de determinar las probabilidades estacionarias que resolver el conjunto de ecuaciones (10.1). Suponga que existen números positivos $x_j, j = 1, \dots, N$ tales que

$$x_i P_{ij} = x_j P_{ji} \quad \text{para } i \neq j, \quad \sum_{j=1}^N x_j = 1$$

Al sumar las ecuaciones anteriores sobre todos los estados i se obtiene

$$\sum_{i=1}^N x_i P_{ij} = x_j \sum_{i=1}^N P_{ji} = x_j$$

y como $\{\pi_j, j = 1, \dots, N\}$ son la única solución de (10.1), esto implica que

$$\pi_j = x_j$$

Cuando $\pi_i P_{ij} = \pi_j P_{ji}$, para toda $i \neq j$, la cadena de Markov es *reversible en el tiempo*, pues se puede mostrar que bajo esta condición, si el estado inicial se elige de acuerdo con las probabilidades $\{\pi_j\}$ y se inicia en cualquier momento la sucesión de estados que va hacia atrás en el tiempo, ésta será también una cadena de Markov con probabilidades de transición P_{ij} .

Suponga ahora que queremos generar el valor de una variable aleatoria X con función de masa de probabilidad $P\{X = j\} = p_j, j = 1, \dots, N$. Si pudiéramos generar una cadena de Markov irreducible aperiódica con probabilidades límites $p_j, j = 1, \dots, N$, entonces podríamos generar de manera aproximada tal variable aleatoria ejecutando

la cadena durante n pasos para obtener el valor de X_n , donde n es grande. Además, si nuestro objetivo era generar muchas variables aleatorias distribuidas de acuerdo con $p_j, j = 1, \dots, N$, para poder estimar $E[h(X)] = \sum_{j=1}^N h(j)p_j$, entonces también podríamos estimar esta cantidad mediante el estimador $\frac{1}{n} \sum_{i=1}^n h(X_i)$. Sin embargo, como los primeros estados de la cadena de Markov pueden tener una fuerte influencia del estado inicial elegido, en la práctica es común desechar los primeros k estados, para algún valor adecuado de k . Es decir, se utiliza el estimador $\frac{1}{n-k} \sum_{i=k+1}^n h(X_i)$. Es difícil saber con exactitud qué tan grande debe ser el valor de k [aunque el lector avanzado puede consultar Aarts y Korst (1989) para obtener algunos resultados útiles al respecto] y por lo común sólo se utiliza la intuición (que en general funciona, pues la convergencia queda garantizada sin importar el valor utilizado).

Una cuestión importante es cómo utilizar la cadena de Markov simulada para estimar el error cuadrático medio del estimador. Es decir, si $\hat{\theta} = \frac{1}{n-k} \sum_{i=k+1}^n h(X_i)$, cómo

$$\text{ECM} = E \left[\left(\hat{\theta} - \sum_{j=1}^N h(j)p_j \right)^2 \right]$$

Una forma es el método de *medias por lotes*, que funciona como sigue. Se separan los $n - k$ estados generados en s lotes de tamaño r , donde $s = (n - k)/r$ es entero, y sea $Y_j, j = 1, \dots, s$ el promedio del j -ésimo lote. Es decir,

$$Y_j = \frac{1}{r} \sum_{i=k+(j-1)r+1}^{k+jr} h(X_i), \quad j = 1, \dots, s$$

Ahora, consideramos las $Y_j, j = 1, \dots, s$ como si fuesen independientes e idénticamente distribuidas con varianza σ^2 y tomamos su varianza muestral $\sum_{j=1}^s (Y_j - \bar{Y})^2 / (s - 1)$ como estimador de σ^2 . La estimación del error cuadrático medio (ECM) es $\hat{\sigma}^2/s$. El valor adecuado de r depende de la cadena de Markov simulada. Mientras más cerca estén las $X_i, i \geq 1$, de ser independientes e idénticamente distribuidas, menor será el valor de r .

En las dos secciones siguientes mostraremos, para un conjunto dado de números positivos $b_j, j = 1, \dots, N$, la forma de construir una cadena de Markov cuyas probabilidades límite sean $\pi_j = b_j / \sum_{i=1}^N b_i$.

10.2 El algoritmo de Hastings-Metropolis

Sean $b(j), j = 1, \dots, m$ números positivos, y sea $B = \sum_{j=1}^m b(j)$. Suponga que m es grande, que B es difícil de calcular y que queremos simular una variable aleatoria

(o una sucesión de variables aleatorias) con función de masa de probabilidad

$$\pi(j) = b(j)/B, \quad j = 1, \dots, m$$

Una forma de simular una sucesión de variables aleatorias cuyas distribuciones convergen a $\pi(j)$, $j = 1, \dots, m$, consiste en determinar una cadena de Markov que sea fácil de simular y cuyas probabilidades límites sean las π_j . El algoritmo de Hastings-Metropolis proporciona un método para realizar esta tarea. Construye una cadena de Markov reversible en el tiempo con las deseadas probabilidades límites, de la manera siguiente.

Sea Q una matriz de probabilidades de transición de Markov irreducible sobre los enteros $1, \dots, m$, donde $q(i, j)$ representa el elemento de Q en el renglón i y la columna j . Definimos una cadena de Markov $\{X_n, n \geq 0\}$ como sigue. Cuando $X_n = i$, se genera una variable aleatoria X tal que $P\{X = j\} = q(i, j)$, $j = 1, \dots, m$. Si $X = j$, entonces X_{n+1} es igual a j con probabilidad $\alpha(i, j)$ y es igual a i con probabilidad $1 - \alpha(i, j)$. En estas condiciones, es fácil ver que la sucesión de estados formará una cadena de Markov con probabilidades de transición P_{ij} dadas por

$$P_{i,j} = q(i, j)\alpha(i, j), \quad \text{si } j \neq i$$

$$P_{i,i} = q(i, i) + \sum_{k \neq i} q(i, k)(1 - \alpha(i, k))$$

Esta cadena de Markov será reversible en el tiempo y tendrá probabilidades estacionarias $\pi(j)$ si

$$\pi(i)P_{ij} = \pi(j)P_{ji} \quad \text{para } j \neq i$$

que es equivalente a

$$\pi(i)q(i, j)\alpha(i, j) = \pi(j)q(j, i)\alpha(j, i)$$

Ahora es fácil verificar que esto se satisface si consideramos

$$\alpha(i, j) = \min\left(\frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}, 1\right) = \min\left(\frac{b(j)q(j, i)}{b(i)q(i, j)}, 1\right)$$

[Para verificar, observe que si $\alpha(i, j) = \pi(j)q(j, i)/\pi(i)q(i, j)$ entonces $\alpha(j, i) = 1$ y viceversa.]

El lector debe observar que el valor de B no es necesario para definir la cadena de Markov, pues bastan los valores $b(j)$. Además, casi siempre ocurre que $\pi(j)$, $j = 1, \dots, m$ no sólo serán probabilidades estacionarias sino también probabilidades límites (de hecho, una condición suficiente es que $P_{i,i} > 0$ para alguna i).

A continuación resumimos el algoritmo de Hastings-Metropolis para generar una cadena de Markov reversible en el tiempo, cuyas probabilidades límites son $\pi(j) = b(j)/B$, $j = 1, \dots, m$.

1. Elegir una matriz Q de probabilidades de transición, de Markov, irreducible, con probabilidades de transición $q(i, j)$, $i, j = 1, \dots, m$. Además, elegir algún valor entero k entre 1 y m .
2. Sean $n = 0$ y $X_0 = k$.
3. Generar una variable aleatoria X tal que $P\{X = j\} = q(X_n, j)$ y generar un número aleatorio U .
4. Si $U < [b(X)q(X, X_n)]/[b(X_n)q(X_n, X)]$, entonces $NS = X$; en caso contrario, $NS = X_n$.
5. $n = n + 1$, $X_n = NS$.
6. Ir al paso 3.

Ejemplo 10a Suponga que queremos generar un elemento aleatorio de un conjunto "combinatorio" complejo y de gran tamaño \mathcal{S} . Por ejemplo, \mathcal{S} podría ser el conjunto de todas las permutaciones (x_1, \dots, x_n) de los números $(1, \dots, n)$ para las cuales $\sum_{j=1}^n jx_j > a$ para una constante a dada; o \mathcal{S} podría ser el conjunto de todas las subgráficas de una gráfica dada, con la propiedad de que para cualquier par de vértices i y j existe un único camino en la subgráfica de i a j (tales subgráficas se llaman árboles).

Para lograr esto, nos valemos del algoritmo de Hastings-Metropolis. Comenzamos suponiendo que se puede definir un concepto de elementos "ceranos" de \mathcal{S} y entonces construiremos una gráfica cuyo conjunto de vértices es \mathcal{S} , colocando un arco entre cada par de elementos cercanos en \mathcal{S} . Por ejemplo, si \mathcal{S} es el conjunto de permutaciones (x_1, \dots, x_n) para las cuales $\sum_{j=1}^n jx_j > a$, entonces podemos decir que dos de estas permutaciones son cercanas si una se obtiene mediante el intercambio de dos posiciones de la otra. Es decir, $(1, 2, 3, 4)$ y $(1, 2, 4, 3)$ son cercanas, pero $(1, 2, 3, 4)$ y $(1, 3, 4, 2)$ no lo son. Si \mathcal{S} es un conjunto de árboles, entonces podemos decir que dos árboles son cercanos si todos los arcos de uno de los árboles, excepto uno, también son arcos del otro árbol.

Ya con este concepto de elementos cercanos, definimos la función de probabilidad de transición q como sigue. Si $N(s)$ es el conjunto de vecinos de s y $|N(s)|$ es igual al número de elementos del conjunto $N(s)$, sea

$$q(s, t) = \frac{1}{|N(s)|}, \quad \text{si } t \in N(\mathcal{S})$$

Es decir, el siguiente estado objetivo de s puede ser cualquiera de sus vecinos, con la misma probabilidad.

Como las probabilidades límites deseadas para la cadena de Markov son $\pi(s) = C$, se tiene que $\pi(s) = \pi(t)$, de modo que

$$\alpha(s, t) = \min(|N(s)|/|N(t)|, 1)$$

Es decir, si el estado actual de la cadena de Markov es s , entonces se elige uno de sus vecinos al azar, digamos t . Si t es un estado con menos vecinos que s (en el lenguaje de la teoría de gráficas, si el grado del vértice t es menor que el del vértice s), entonces el estado siguiente es t . En caso contrario, se genera un número aleatorio U , y el siguiente estado es t si $U < |N(s)|/|N(t)|$, y es igual a s en caso contrario. Las probabilidades límites de esta cadena de Markov son $\pi(s) = 1/|\mathcal{S}|$. ■

Ejemplo 10b En el ejemplo 10a, suponga que queremos generar un elemento elegido al azar en el conjunto de todas las permutaciones $x = (x_1, \dots, x_n) \in \mathcal{S}$ donde $\mathcal{S} = x: \sum_{j=1}^n jx_j > a$. Al aplicar el método detallado en el ejemplo 10a, comenzando con x , ¿cómo elegimos de manera aleatoria una permutación cercana a \mathcal{S} ? Bueno, si n es pequeño, entonces es fácil llevar un registro de todos los vecinos que están en \mathcal{S} y elegir uno al azar, llamado y , como el siguiente estado objetivo. Habría que determinar entonces el número de vecinos de y , y entonces el estado siguiente de la cadena de Markov sería y con probabilidad $\min(1, |N(x)|/|N(y)|)$, o seguiría siendo x . Sin embargo, si n es grande, esto puede no ser práctico; un método mejor consistiría en desarrollar el espacio de estados, el cual consta de todas las $n!$ permutaciones. La función de masa de probabilidad límite deseada es entonces

$$\begin{aligned} \pi(x) &= C, & x \in \mathcal{S} \\ \pi(x) &= 0, & x \notin \mathcal{S} \end{aligned}$$

En este marco, cada permutación x tiene vecinos y cualquiera se puede elegir en forma aleatoria, generando un subconjunto aleatorio de tamaño dos a partir del conjunto $1, \dots, n$; si se eligen i y j , entonces el siguiente estado objetivo y se obtiene intercambiando los valores de x_i y x_j . Si $y \in \mathcal{S}$, entonces y se convierte en el estado siguiente de la cadena y , en caso contrario, entonces el estado siguiente seguiría siendo x . ■

10.3 El muestreador de Gibbs

La versión más utilizada del algoritmo de Hastings-Metropolis es el muestreador de Gibbs. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con función de masa de probabilidad (o función de densidad de probabilidad en el caso continuo) $p(x)$, la cual sólo está determinada salvo una constante multiplicativa, y suponga que queremos generar un

vector aleatorio cuya distribución sea la de la distribución condicional de X , dado que $X \in A$ para algún conjunto A . Es decir, queremos generar un vector aleatorio con función de masa

$$f(x) = \frac{p(x)}{P\{X \in A\}}, \quad \text{para } x \in A$$

El muestreador de Gibbs supone que para cualquier $i, i = 1, \dots, n$ y cualesquiera valores $x_j, j \neq i$, podemos generar una variable aleatoria X con la función de masa de probabilidad

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$$

Esto opera considerando una cadena de Markov con estados $x = (x_1, \dots, x_i, \dots, x_n) \in A$; luego se utiliza el algoritmo de Hastings-Metropolis con las probabilidades de transición de Markov definidas como sigue. Siempre que el estado actual sea x , se genera una coordenada que tiene la misma probabilidad de ser cualquiera de los valores $1, \dots, n$. Si la coordenada i es la elegida, entonces se genera una variable aleatoria X con función de masa de probabilidad $P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$ y si $X = x$, entonces se considera el estado $y = (x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$ para la transición. En otras palabras, el muestreador de Gibbs emplea el algoritmo de Hastings-Metropolis con

$$q(x, y) = \frac{1}{n} P\{X_i = x | X_j = x_j, j \neq i\} = \frac{1}{n} \frac{p(y)}{P\{X_j = x_j, j \neq i\}}$$

Como la función de masa objetivo es f , el vector y se acepta como el nuevo estado con probabilidad

$$\alpha(x, y) = \min\left(\frac{f(y)q(y, x)}{f(x)q(x, y)}, 1\right)$$

Como para $x \in A, y \in A$, se tiene que

$$\frac{f(y)q(y, x)}{f(x)q(x, y)} = \frac{f(y)p(x)}{f(x)p(y)} = 1$$

mientras que para $x \in A$ y $y \notin A$

$$\frac{f(y)q(y, x)}{f(x)q(x, y)} = 0$$

vemos que el siguiente estado es y si $y \in A$, o sigue siendo x si $y \notin A$.

En resumen, vemos que la cadena de Markov reversible en el tiempo, con probabilidades estacionarias dadas por f , generada por el muestreador de Gibbs es la siguiente.

1. Sea $x = (x_1, \dots, x_n)$ un vector en A para el cual $p(x) > 0$.
2. Sea I una variable aleatoria que toma uno de los valores $1, \dots, n$, donde cada valor tiene la misma probabilidad de ocurrir.

3. Si $I = i$, generar el valor de una variable aleatoria X tal que

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$$

4. Si $X = x$ y $(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \in A$, entonces el nuevo valor de x_i es igual a x . En caso contrario, se mantiene el valor de x_i .
5. Regresar al paso 2.

Así, en cada paso se elige una de las variables X_i al azar y se genera una variable aleatoria con la distribución condicional de X_i dado que $X_j = x_j, j \neq i$. Si el nuevo vector, donde este nuevo valor reemplaza a x_i , está en A , entonces es el siguiente estado de la cadena; si el vector no está en A , entonces el estado permanece sin cambios.

Ejemplo 10c Suponga que queremos generar n puntos aleatorios en el círculo de radio 1 con centro en el origen, condicionando sobre el evento de que no haya dos puntos a menos de una distancia d entre ellos, donde

$$\beta = P\{\text{no haya dos puntos a menos de una distancia } d \text{ entre ellos}\}$$

es un número positivo pequeño (si β no fuese pequeño, entonces podríamos continuar generando conjuntos de n puntos aleatorios en el círculo, deteniéndonos la primera vez que no haya dos puntos a menos de una distancia d entre ellos). Esto se puede lograr mediante el muestreador de Gibbs, comenzando con n puntos del círculo, x_1, \dots, x_n , tales que no haya dos puntos a menos de una distancia d entre ellos. Entonces generamos un número aleatorio U y hacemos $I = \text{Ent}(nU) + 1$. Además, generamos un punto aleatorio en el círculo. Si este punto no está a menos de una distancia d de cualquiera de los demás $n - 1$ puntos, excluyendo a x_I , entonces reemplazamos a x_I por este punto generado; en caso contrario, no hacemos la sustitución. Después de un gran número de iteraciones, el conjunto de n puntos tendrá aproximadamente la distribución deseada. ■

En el siguiente ejemplo, el objetivo es generar X , sin la restricción de que deba pertenecer a algún conjunto de valores (es decir, A se considera como todos los valores de X).

Ejemplo 10d Redes de colas. Suponga que r individuos se mueven entre $m + 1$ estaciones de colas, y sea $X_i(t)$, $i = 1, \dots, m$, el número de individuos en la estación i en el instante t . Si

$$p(n_1, \dots, n_m) = \lim_{t \rightarrow \infty} P\{X_i(t) = n_i, i = 1, \dots, m\}$$

entonces, si suponemos que los tiempos de servicio se distribuyen en forma exponencial, con frecuencia se puede establecer que

$$p(n_1, \dots, n_m) = C \prod_{i=1}^m P_i(n_i), \quad \text{si } \sum_{i=1}^m n_i \leq r$$

$$p(n_1, \dots, n_m) = 0, \quad \text{en caso contrario}$$

donde $P_i(n)$, $n \geq 0$ es una función de masa de probabilidad para cada $i = 1, \dots, m$. Tal función de masa de probabilidad conjunta tiene una *forma producto*.

Aunque es relativamente directo establecer que $p(n_1, \dots, n_m)$ tiene la forma producto anterior y determinar las funciones de masa P_i , puede ser difícil calcular la constante C . Ya que aunque

$$C \sum_{n: s(n) \leq r} \prod_{i=1}^m P_i(n_i) = 1$$

donde $n = (n_1, \dots, n_m)$ y $s(n) = \sum_{i=1}^m n_i$, puede ser difícil utilizar este resultado. Esto se debe a que la suma es sobre todos los vectores enteros no negativos n para los cuales $\sum_{i=1}^m n_i \leq r$ y existen $\binom{r+m}{m}$ de tales vectores, que es un número relativamente grande, aunque m y r sean de tamaño moderado.

Otro método para averiguar el comportamiento de $p(n_1, \dots, n_m)$, que burla las dificultades computacionales del cálculo de C , consiste en usar el muestreador de Gibbs para generar una sucesión de valores con una distribución aproximadamente igual a la de p .

Para comenzar, observemos que si $N = (N_1, \dots, N_m)$ tiene la función de masa conjunta p , entonces para $n = 0, \dots, r - \sum_{k \neq i} n_k$

$$\begin{aligned} P\{N_i = n | N_1 = n_1, \dots, N_{i-1} = n_{i-1}, N_{i+1} = n_{i+1}, \dots, N_m = n_m\} \\ = \frac{p(n_1, \dots, n_{i-1}, n, n_{i+1}, \dots, n_m)}{\sum_j p(n_1, \dots, n_{i-1}, j, n_{i+1}, \dots, n_m)} \\ = \frac{P_i(n)}{\sum_j P_i(j)} \end{aligned}$$

donde la suma anterior es sobre toda $j = 0, \dots, r - \sum_{k \neq i} n_k$. En otras palabras, la distribución condicional de N_i dados los valores de $N_j, j \neq i$, es igual a la distribución condicional de una variable aleatoria con función de masa P_i dado que su valor es menor o igual a $r - \sum_{j \neq i} N_j$.

Así, podemos generar los valores de una cadena de Markov cuya función de masa de probabilidad límite es $p(n_1, \dots, n_m)$ como sigue:

1. Sean (n_1, \dots, n_m) enteros no negativos arbitrarios que satisfacen $\sum_i n_i \leq r$.
2. Generar U y sea $I = \text{Ent}(mU) + 1$.
3. Si $I = i$, sea X_i con función de masa P_i y generar una variable aleatoria N cuya distribución es la distribución condicional de X_i dado que $X_i \leq r - \sum_{j \neq i} n_j$.
4. Sea $n_i = N$ e ir al paso 2.

Los valores sucesivos de (n_1, \dots, n_m) constituyen la sucesión de estados de una cadena de Markov con la distribución límite p . Todas las cantidades de interés relacionadas con p se pueden estimar a partir de esta sucesión. Por ejemplo, el promedio de los valores de la j -ésima coordenada de estos vectores convergerá en el número medio de individuos en la estación j , la proporción de vectores cuya j -ésima coordenada es menor que k convergerá en la probabilidad límite de que el número de individuos en la estación j sea menor que k , etcétera. ■

En los casos en que es fácil generar X_i condicionando sobre los valores de X_j , $j \neq i$ y sobre el hecho $X \in A$, tal generación puede reemplazar los pasos 3 y 4 en el muestreador de Gibbs. Esto se ilustra en el siguiente ejemplo.

Ejemplo 10e Sean X_i , $i = 1, \dots, n$, variables aleatorias independientes, donde X_i tiene una distribución exponencial con razón λ_i , $i = 1, \dots, n$. Sea $S = \sum_{i=1}^n X_i$ y suponga que queremos generar el vector aleatorio $X = (X_1, \dots, X_n)$ condicionando sobre el evento de que $S > c$ para alguna constante positiva grande c . Es decir, queremos generar el valor de un vector aleatorio cuya función de densidad está dada por

$$f(x_1, \dots, x_n) = \frac{1}{P\{S > c\}} \prod_{i=1}^n \lambda_i e^{-\lambda_i x_i}, \quad \text{si } \sum_{i=1}^n x_i > c$$

Esto se logra fácilmente comenzando con un vector inicial $x = (x_1, \dots, x_n)$ que satisface $x_i > 0$, $i = 1, \dots, n$, y $\sum_{i=1}^n x_i > c$. Luego generamos un número aleatorio U y hacemos $I = \text{Ent}(nU) + 1$. Supongamos que $I = i$. Queremos generar una variable aleatoria exponencial X con razón λ_i , condicionada sobre el evento de que $X + \sum_{j \neq i} x_j > c$. Es decir, queremos generar el valor de X condicionando sobre el evento de que excede a $c - \sum_{j \neq i} x_j$. Por lo tanto, utilizando el hecho de que una exponencial condicionada a ser mayor que una constante positiva se distribuye como la constante más la exponencial, vemos que debemos generar una variable aleatoria exponencial Y con razón λ_i [digamos que $Y = -1/\lambda_i \log U$], y hacemos

$$X = Y + \left(c - \sum_{j \neq i} x_j \right)^+$$

donde b^+ es igual a b cuando $b > 0$ y es igual a 0 en caso contrario. El valor de x_i debe igualarse a X e iniciar otra iteración del algoritmo. ■

Suponga que estamos interesados en estimar

$$\alpha = P\{h(X) > a\}$$

donde $X = (X_1, \dots, X_n)$ es un vector aleatorio que se puede generar mediante un muestreo de Gibbs, h es una función arbitraria de X y α es muy pequeña. Como un valor generado de $h(X)$ siempre será menor que a , se necesitaría mucho tiempo para obtener un estimador cuyo error sea pequeño con respecto de α si utilizamos un método directo del muestreador de Gibbs para generar una sucesión de vectores aleatorios cuya distribución converge a la de X . Sin embargo, consideremos el siguiente método.

Para comenzar, observe que para valores $-\infty = a_0 < a_1 < a_2 < \dots < a_k = a$,

$$\alpha = \prod_{i=1}^k P\{h(X) > a_i | h(X) > a_{i-1}\}$$

Así, podemos obtener un estimador de α considerando el producto de estimadores de las cantidades $P\{h(X) > a_i | h(X) > a_{i-1}\}$, para $i = 1, \dots, k$. Para que esto sea eficiente, debemos elegir los valores a_i , $i = 1, \dots, k$ de modo que $P\{h(X) > a_i | h(X) > a_{i-1}\}$ sean todos de tamaño moderado.

Para estimar $P\{h(X) > a_i | h(X) > a_{i-1}\}$, utilizamos el muestreador de Gibbs como sigue.

1. Hacer $J = 0$.
2. Elegir un vector x tal que $h(x) > a_{i-1}$.
3. Generar un número aleatorio U y hacer $I = \text{Ent}(nU) + 1$.
4. Si $I = k$, generar X con la distribución condicional de X_k , dado que $X_j = x_j$, $j \neq k$.
5. Si $h(x_1, \dots, x_{k-1}, X, x_{k+1}, \dots, x_n) > a_{i-1}$, hacer $x_k = X$.
6. Si $h(x_1, \dots, x_k, \dots, x_n) > a_i$, hacer $J = J + 1$.
7. Ir al paso 3.

El cociente del valor final de J entre el número de iteraciones de este algoritmo es el estimador de $P\{h(X) > a_i | h(X) > a_{i-1}\}$.

Ejemplo 10f Suponga en el modelo de la red de colas del ejemplo 10d que los tiempos de servicio en el servidor i son exponenciales con razón μ_i , $i = 1, \dots, m + 1$ y que cuando un cliente concluye el servicio en el servidor i , entonces, independientemente

de todo lo demás, ese cliente se mueve para formarse en la fila (o entra a servicio si el servidor está libre) del servidor j con probabilidad P_{ij} , donde $\sum_{j=1}^{m+1} P_{ij} = 1$. Se puede mostrar que la función de masa de probabilidad límite del número de clientes en los servidores $1, \dots, m$ está dado, para $\sum_{j=1}^m n_j \leq r$, por

$$p(n_1, \dots, n_m) = C \prod_{j=1}^m \left(\frac{\pi_j \mu_{m+1}}{\pi_{m+1} \mu_j} \right)^{n_j}$$

donde π_j , $j = 1, \dots, m+1$, son las probabilidades estacionarias de la cadena de Markov con probabilidades de transición P_{ij} . Es decir, son las únicas soluciones de

$$\pi_j = \sum_{i=1}^{m+1} \pi_i P_{ij}$$

$$\sum_{j=1}^{m+1} \pi_j = 1$$

Si numeramos los servidores de modo que $\max(\pi_j/\mu_j) = \pi_{m+1}/\mu_{m+1}$, entonces al hacer $a_j = \pi_j \mu_{m+1} / \pi_{m+1} \mu_j$, tenemos que para $\sum_{j=1}^m n_j \leq r$.

$$p(n_1, \dots, n_m) = C \prod_{j=1}^m (a_j)^{n_j}$$

donde $0 \leq a_j \leq 1$. Esto implica fácilmente que la distribución condicional del número de clientes en el servidor i , dados los números n_j , $j \neq i$, en los otros $m-1$ servidores, es como la distribución condicional de -1 más una variable aleatoria geométrica con parámetro $1 - a_i$, dado que la geométrica es menor o igual a $r+1 - \sum_{j \neq i} n_j$.

Cuando π_j y μ_j son constantes para toda j , la distribución condicional del número de clientes en el servidor i , dados los números n_j , $j \neq i$, en los otros servidores excepto el servidor $m+1$, es la distribución discreta uniforme en $0, 1, \dots, r - \sum_{j \neq i} n_j$. Suponga que éste es el caso y que $m=20$, $r=100$, y que estamos interesados en estimar la probabilidad límite de que el número de clientes en el servidor 1 (denotado X_1) es mayor que 18. Si $t_0 = -1$, $t_1 = 5$, $t_2 = 9$, $t_3 = 12$, $t_4 = 15$, $t_5 = 17$, $t_6 = 18$, podemos emplear el muestreador de Gibbs para estimar de manera sucesiva las cantidades $P\{X_1 > t_i | X_1 > t_{i-1}\}$, $i = 1, 2, 3, 4, 5, 6$. Por ejemplo, estimamos $P\{X_1 > 17 | X_1 > 15\}$ comenzando con un vector n_1, \dots, n_{20} para el cual $n_1 > 15$ y $s = \sum_{i=1}^{20} n_i \leq 100$. Luego generamos un número aleatorio U y hacemos $I = \text{Ent}(20U + 1)$. Ahora generamos un segundo número aleatorio V y hacemos $J = \text{Ent}((100 + n_I - s)V + 1)$. Si $I \neq 1$, entonces n_1 se iguala a J . Si $I = 1$ y $J > 15$, entonces n_1 es igual a J ; en caso contrario, su valor o se modifica y comienza la siguiente itera-

ción del algoritmo. La fracción de iteraciones para las cuales $n_1 > 17$ es la estimación de $P\{X_1 > 17 | X_1 > 15\}$. ■

La idea de escribir una pequeña probabilidad como el producto de probabilidades condicionales de tamaño moderado y luego estimar cada una de las probabilidades condicionales no necesita el muestreador de Gibbs. Otra variante del algoritmo de Hastings-Metropolis podría ser más adecuada. Ilustraremos esto mediante un ejemplo que ya ha sido estudiado con muestreo de importancia, el ejemplo 8u.

Ejemplo 10g Suponga que estamos interesados en estimar el número de permutaciones $x = (x_1, \dots, x_n)$ para las cuales $t(x) > a$, donde $t(x) = \sum_{j=1}^n jx_j$ y a es tal que este número de permutaciones es mucho menor que $n!$. Si $X = (X_1, \dots, X_n)$ es igual a cualquiera de las $n!$ permutaciones con la misma probabilidad, y si

$$\alpha = P\{T(X) > a\}$$

entonces α es pequeña y la cantidad de interés es $\alpha n!$. Si $0 = a_0 < a_1 < \dots < a_k = a$, tenemos que

$$\alpha = \prod_{i=1}^k P\{T(X) > a_i | T(X) > a_{i-1}\}$$

Para estimar $P\{T(X) > a_i | T(X) > a_{i-1}\}$ utilizamos el algoritmo de Hastings-Metropolis como en los ejemplos 10a o 10b para generar una cadena de Markov cuya distribución límite es

$$\pi(x) = \frac{1}{N_{i-1}}, \quad \text{si } T(x) > a_{i-1}$$

donde N_{i-1} es el número de permutaciones x tales que $T(x) > a_{i-1}$. La proporción de los estados generados x de esta cadena de Markov que tienen $T(x) > a_i$ es la estimación de $P\{T(X) > a_i | T(X) > a_{i-1}\}$. ■

En muchas aplicaciones, es relativamente sencillo reconocer la forma de las distribuciones condicionales necesarias en el muestreador de Gibbs.

Ejemplo 10h Suponga que para cierta función no negativa $h(y, z)$, la densidad conjunta de las variables aleatorias no negativas X, Y y Z es

$$f(x, y, z) = Cx^{y-1}(1-x)^y h(y, z), \quad \text{para } 0 < x < 0.5$$

Entonces la densidad condicional de X dado que $Y = y$ y $Z = z$ es

$$f(x|y, z) = \frac{f(x, y, z)}{f_{YZ}(y, z)}$$

Como y y z son fijos y x es el argumento de esta densidad condicional, podemos escribir lo anterior como

$$f(x|y, z) = C_1 f(x, y, z)$$

donde C_1 no depende de x . Por lo tanto, tenemos que

$$f(x|y, z) = C_2 x^{y-1} (1-x)^{zy}, \quad 0 < x < 0.5$$

donde C_2 no depende de x . Pero podemos reconocer esto como la densidad condicional de una variable aleatoria beta con parámetros y y $zy + 1$ que está condicionada a estar en el intervalo $(0, 0.5)$. ■

En vez de elegir siempre una coordenada aleatoria para actualizarla, el muestreador de Gibbs también puede considerar las coordenadas en orden. Es decir, en la primera iteración podríamos hacer $I = 1$, luego hacer $I = 2$ en la siguiente iteración, luego $I = 3$, y así hasta la n -ésima iteración, donde $I = n$. En la siguiente iteración comenzamos de nuevo. Ilustramos esto con nuestro siguiente ejemplo, relacionado con el modelado de cuadrangulares logrados por dos de los mejores bateadores del béisbol profesional.

Ejemplo 10i Sea $N_1(t)$ el número de cuadrangulares logrados en el primer $100t$ por ciento de una temporada de béisbol, $0 \leq t \leq 1$, por Matt Williams, jugador de los Gigantes de San Francisco; de manera análoga, sea $N_2(t)$ el número logrado por Frank Thomas, jugador de los Medias Blancas de Chicago.

Suponga que existen variables aleatorias W_1 y W_2 tales que dado que $W_1 = w_1$ y $W_2 = w_2$, $\{N_1(t), 0 \leq t \leq 1\}$ y $\{N_2(t), 0 \leq t \leq 1\}$ son procesos Poisson independientes con razones w_1 y w_2 , respectivamente. Además, suponga que W_1 y W_2 son variables aleatorias exponenciales independientes con razón Y , que a su vez es una variable aleatoria con distribución uniforme entre 0.02 y 0.10. En otras palabras, la hipótesis es que el jugador batea un cuadrangular de acuerdo con un proceso Poisson cuyas razones son variables aleatorias de una distribución definida en términos de un parámetro, que a su vez es una variable aleatoria con una distribución dada.

Suponga que Williams ha bateado 25 cuadrangulares y Thomas 18 en la primera mitad de la temporada. Dé un método para estimar el número medio de cuadrangulares que batea cada uno en toda la temporada.

Solución Para resumir el modelo, existen variables aleatorias Y, W_1, W_2 tales que:

1. Y es uniforme en $(0.02, 0.10)$.
2. Dado que $Y = y$, W_1 y W_2 son variables aleatorias exponenciales, independientes e idénticamente distribuidas con razón y .
3. Dado que $W_1 = w_1$ y $W_2 = w_2$, $\{N_1(t)\}$ y $\{N_2(t)\}$ son procesos Poisson independientes con razones w_1 y w_2 .

Para determinar $E[N_1(1)|N_1(0.5) = 25, N_2(0.5) = 18]$, primero condicionamos sobre W_1 .

$$E[N_1(1)|N_1(0.5) = 25, N_2(0.5) = 18, W_1] = 25 + 0.5W_1$$

Al calcular la esperanza condicional, dado que $N_1(0.5) = 25$ y $N_2(0.5) = 18$, obtenemos de lo anterior que

$$E[N_1(1)|N_1(0.5) = 25, N_2(0.5) = 18] = 25 + 0.5E[W_1|N_1(0.5) = 25, N_2(0.5) = 18]$$

De manera análoga,

$$E[N_2(1)|N_1(0.5) = 25, N_2(0.5) = 18] = 18 + 0.5E[W_2|N_1(0.5) = 25, N_2(0.5) = 18]$$

Podemos estimar estas esperanzas condicionales mediante el muestreador de Gibbs. Para comenzar, analizamos la distribución conjunta: para $0.02 < y < 0.10$, $w_1 > 0$, $w_2 > 0$,

$$f(y, w_1, w_2, N_1(0.5) = 25, N_2(0.5) = 18) = Cy^2 e^{-(w_1 + w_2)y} e^{-(w_1 + w_2)/2} (w_1)^{25} (w_2)^{18}$$

donde C no depende de y, w_1 o w_2 . Por lo tanto, para $0.02 < y < 0.10$,

$$f(y|w_1, w_2, N_1 = 25, N_2 = 18) = C_1 y^2 e^{-(w_1 + w_2)y}$$

lo cual muestra que la distribución condicional de Y dados $w_1, w_2, N_1 = 25, N_2 = 18$, es la de una variable aleatoria gamma con parámetros 3 y $w_1 + w_2$ condicionada a estar entre 0.02 y 0.10. Además,

$$f(w_1|y, w_2, N_1(0.5) = 25, N_2(0.5) = 18) = C_2 e^{-(y+1/2)w_1} (w_1)^{25}$$

de lo cual podemos concluir que la distribución condicional de W_1 dados $y, w_2, N_1 = 25, N_2 = 18$ es gamma con parámetros 26 y $y + \frac{1}{2}$. De manera análoga, la distribución condicional de W_2 dados $y, w_1, N_1 = 25, N_2 = 18$ es gamma con parámetros 19 y $y + \frac{1}{2}$.

Por lo tanto, si comenzamos con los valores y, w_1, w_2 , donde $0.02 < y < 0.10$ y $w_1 > 0$, el muestreador de Gibbs es como sigue.

1. Generar el valor de una variable aleatoria gamma con parámetros 3 y $w_1 + w_2$ que está condicionada a estar entre 0.02 y 0.10 y sea éste el nuevo valor de y .
2. Generar el valor de una variable aleatoria gamma con parámetros 26 y $y + \frac{1}{2}$, y sea éste el nuevo valor de w_1 .
3. Generar el valor de una variable aleatoria gamma con parámetros 19 y $y + \frac{1}{2}$, y sea éste el nuevo valor de w_2 .
4. Regresar al paso 1.

El promedio de los valores de w_1 es nuestra estimación de $E[W_1 | N_1(0.5) = 25, N_2(0.5) = 18]$, y el promedio de los valores de w_2 es nuestra estimación de $E[W_2 | N_1(0.5) = 25, N_2(0.5) = 18]$. La mitad del primero más 25 es nuestra estimación del número medio de cuadrangulares que Matt Williams bateará durante un año, y la mitad del segundo más 18 es nuestra estimación del número medio de los que bateará Frank Thomas.

Debemos observar que los números de cuadrangulares bateados por los dos jugadores son dependientes, a causa de su dependencia común del valor de la variable aleatoria Y . Es decir, el valor de Y (que se podría relacionar con cantidades como el grado promedio de viveza de las pelotas de béisbol utilizadas esa temporada o las condiciones climáticas promedio de ese año) influye en la distribución del número medio de cuadrangulares que cada jugador bateará durante el año. Así, la información acerca del número de cuadrangulares bateados por uno de los jugadores proporciona cierta información probabilística acerca del valor de Y , el cual afecta la distribución del número bateado por el otro. Este modelo, en el que hay una variable aleatoria común (Y en este caso) que influye en las distribuciones de los parámetros condicionales de las variables aleatorias de interés, se conoce como modelo *jerárquico de Bayes*. ■

Al aplicar el muestreador de Gibbs, no es necesario condicionar todas las variables excepto una. Si se pueden generar distribuciones condicionales conjuntas, entonces podemos utilizarlas. Por ejemplo, supongamos que $n = 3$ y que podemos generar a partir de la distribución condicional de cualesquiera dos de ellas, dada la tercera. Entonces, en cada iteración, podríamos generar un número aleatorio U , hacer $I = \text{Ent}(3U + 1)$ y generar a partir de la distribución conjunta de las $X_j, X_k, j, k \neq I$, dado el valor actual de X_I . Es decir, si $I = 2$, generaríamos a partir de la distribución conjunta de X_1, X_3 dado el valor actual de X_2 , digamos, $X_2 = x_2$. Si esto produce los valores y_1, y_3 ,

entonces si el vector (y_1, x_2, y_3) está en el conjunto A , se convierte en el nuevo vector de estado; en caso contrario, el estado no se modifica.

Ejemplo 10j Sean $X_i, i = 1, 2, 3, 4, 5$, variables aleatorias exponenciales independientes, donde X_i tiene media i , y suponga que estamos interesados en utilizar simulación para estimar

$$\beta = P\left\{\prod_{i=1}^5 X_i > 120 \mid \sum_{i=1}^5 X_i = 15\right\}.$$

Esto se logra mediante el muestreador de Gibbs, eligiendo en forma aleatoria dos de las coordenadas. Para comenzar, supongamos que X y Y son exponenciales independientes con razones respectivas λ y μ , donde $\mu < \lambda$, y determinemos la distribución condicional de X dado que $X + Y = a$, como sigue.

$$\begin{aligned} f_{X|X+Y}(x|a) &= C_1 f_{X,Y}(x, a-x), & 0 < x < a \\ &= C_2 e^{-\lambda x} e^{-\mu(a-x)}, & 0 < x < a \\ &= C_3 e^{-(\lambda-\mu)x}, & 0 < x < a \end{aligned}$$

lo cual muestra que la distribución condicional es la de una exponencial con razón $\lambda - \mu$ condicionada a ser menor que a .

Con este resultado, podemos estimar β haciendo que el estado inicial $(x_1, x_2, x_3, x_4, x_5)$ sean cinco números positivos arbitrarios que sumen 15. Ahora elegimos al azar dos elementos del conjunto 1, 2, 3, 4, 5; digamos que se han elegido $I = 2$ y $J = 5$. Entonces la distribución condicional de X_2, X_5 dados los otros valores es la distribución condicional de dos exponenciales independientes con medias 2 y 5, dado que su suma es $15 - x_1 - x_3 - x_4$. Pero por lo anterior, podemos obtener los valores de X_2 y X_5 generando el valor de una exponencial con razón $\frac{1}{2} - \frac{1}{5} = \frac{3}{10}$ condicionada a ser menor que $15 - x_1 - x_3 - x_4$, y luego x_2 será igual a ese valor y damos a x_5 un valor para hacer que $\sum_{i=1}^5 x_i = 15$. Este proceso se puede repetir en forma continua, y la fracción de vectores de estado \mathbf{x} tales que $\prod_{i=1}^5 x_i > 120$ es la estimación de β . ■

Observaciones

1. El mismo argumento permite verificar que obtenemos la función de masa límite adecuada cuando consideramos las coordenadas en orden y aplicamos el muestreador de Gibbs (como en el ejemplo 10i), o cuando lo empleamos condicionando sobre todos los valores excepto uno (como en el ejemplo 10j). Estos resultados se demuestran observando que si uno elige el estado inicial de acuerdo con la función de masa f , entonces, en cualquier caso, el estado siguiente

también tiene función de masa f . Pero esto muestra que f satisface las ecuaciones (10.1), y por unicidad f es la función de masa límite.

- Suponga que utiliza el muestreador de Gibbs para estimar $E[X_i]$ en una situación en la que las medias condicionales $E[X_i | X_j, j \neq i]$ se calculan fácilmente. Entonces, en vez de utilizar el promedio de los valores sucesivos de X_i como estimador, por lo general es mejor utilizar el promedio de las esperanzas condicionales. Es decir, si el estado actual es x , entonces se considera $E[X_i | X_j, j \neq i]$ en vez de x_i como el estimador de esa iteración. De manera análoga, si se intenta estimar $P\{X_i = x\}$ y $P\{X_i = x | X_j, j \neq i\}$ se calcula fácilmente, entonces por lo general el promedio de estas cantidades es un mejor estimador que la fracción de tiempo durante el cual el i -ésimo componente del vector de estado es igual a x .
- El muestreador de Gibbs muestra que el conocimiento de todas las distribuciones condicionales de X_i dados los valores de las demás $X_j, j \neq i$, determina la distribución conjunta de X .

10.4 Temple simulado

Sea \mathcal{A} un conjunto finito de vectores y sea $V(x)$ una función no negativa definida en $x \in \mathcal{A}$, y suponga que estamos interesados en determinar su valor máximo y al menos un argumento en el cual se alcanza el valor máximo. Es decir, si

$$V^* = \max_{x \in \mathcal{A}} V(x)$$

y

$$\mathcal{M} = \{x \in \mathcal{A} : V(x) = V^*\}$$

estamos interesados en determinar V^* así como un elemento en \mathcal{M} . Ahora mostraremos cómo lograr esto mediante los métodos de este capítulo.

Para comenzar, sea $\lambda > 0$ y consideremos la siguiente función de masa de probabilidad sobre el conjunto de valores de \mathcal{A} :

$$p_\lambda(x) = \frac{e^{\lambda V(x)}}{\sum_{x \in \mathcal{A}} e^{\lambda V(x)}}$$

Al multiplicar el numerador y el denominador de lo anterior por $e^{-\lambda V^*}$ y denotando como $|\mathcal{M}|$ al número de elementos en \mathcal{M} , vemos que

$$p_\lambda(x) = \frac{e^{\lambda(V(x) - V^*)}}{|\mathcal{M}| + \sum_{x \notin \mathcal{M}} e^{\lambda(V(x) - V^*)}}$$

Sin embargo, como $V(x) - V^* < 0$ para $x \notin \mathcal{M}$, obtenemos que cuando $\lambda \rightarrow \infty$,

$$p_\lambda(x) \rightarrow \frac{\delta(x, \mathcal{M})}{|\mathcal{M}|}$$

donde $\delta(x, \mathcal{M}) = 1$ si $x \in \mathcal{M}$ y es igual a cero en caso contrario.

Por lo tanto, si λ es grande y generamos una cadena de Markov cuya distribución límite es $p_\lambda(x)$, entonces la mayor parte de esta distribución límite se concentrará en los puntos de \mathcal{M} . Un método que con frecuencia es útil al definir tal cadena es introducir el concepto de vectores cercanos y luego utilizar un algoritmo de Hastings-Metropolis. Por ejemplo, podríamos decir que los dos vectores $x \in \mathcal{A}$ y $y \in \mathcal{A}$ son cercanos si difieren sólo en una coordenada o si uno se puede obtener del otro intercambiando dos de sus componentes. Entonces podríamos hacer que el estado siguiente objetivo desde x sea cualquiera de sus vectores cercanos, todos con la misma probabilidad, y si se elige al vector cercano y , entonces el estado siguiente se convierte en y con probabilidad

$$\min \left\{ 1, \frac{e^{\lambda V(y)} / |N(y)|}{e^{\lambda V(x)} / |N(x)|} \right\}$$

o sigue siendo x en caso contrario, donde $|N(z)|$ es el número de vecinos de z . Si cada vector tiene el mismo número de vecinos (y si no es así, siempre podemos lograrlo incrementando el espacio de estados y haciendo que el valor V de cualquier nuevo estado sea igual a 0); luego, cuando el estado es x , se elige uno de sus vecinos, digamos y ; y si $V(y) \geq V(x)$, entonces la cadena se mueve al estado y , y si $V(y) < V(x)$, entonces la cadena se mueve al estado y con probabilidad $\exp\{\lambda(V(y) - V(x))\}$ o permanece en el estado x en caso contrario.

Una limitación del algoritmo anterior es que como λ se eligió de gran tamaño, si la cadena entra a un estado x cuyo valor V sea mayor que el de cada uno de sus vecinos, entonces podría tardar mucho tiempo en moverse a un estado diferente. Es decir, aunque se necesita un valor grande de λ para que la distribución límite ponga la mayor parte de su peso en los puntos de \mathcal{M} , por lo general tal valor requiere un número muy grande de transiciones antes de tender a la distribución límite. La segunda limitación es que, como sólo hay un número finito de valores posibles de x , todo el concepto de convergencia pierde sentido, pues siempre podríamos, en teoría, probar con cada uno de los valores posibles y así obtener la convergencia en un número finito de pasos. Así, en vez de considerar lo anterior desde un punto de vista estrictamente matemático, tiene más sentido utilizarlo como un método heurístico, y al hacer esto se ha visto que es útil permitir que el valor de λ cambie con el tiempo.

Una variante popular de lo anterior, el *temple simulado*, opera como sigue. Si el n -ésimo estado de la cadena de Markov es \mathbf{x} , entonces se elige al azar un valor cercano. Si éste es \mathbf{y} , entonces el estado siguiente es \mathbf{y} con probabilidad

$$\min \left\{ 1, \frac{e^{\lambda_n V(\mathbf{y})} / |N(\mathbf{y})|}{e^{\lambda_n V(\mathbf{x})} / |N(\mathbf{x})|} \right\}$$

o sigue siendo \mathbf{x} , donde λ_n , $n \geq 1$, es un conjunto prescrito de valores que comienzan siendo pequeños (lo cual produce un gran número de cambios de estado) y luego crecen.

Una elección útil (desde el punto de vista computacional) de λ_n (y una elección que matemáticamente produce la convergencia) consiste en hacer $\lambda_n = C \log(1 + n)$, donde $C > 0$ es cualquier constante positiva fija (véase Besag *et al.*, 1995; Diaconis y Holmes, 1995). Si generamos m estados sucesivos X_1, \dots, X_m , podemos estimar V^* mediante $\max_{i=1, \dots, m} V(X_i)$, y si el máximo ocurre en X_{i^*} , entonces éste se considera como un punto estimado en \mathcal{M} .

Ejemplo 10k El problema del agente de ventas viajero. Una versión del problema del agente de ventas viajero es que éste comience en la ciudad 0 y luego visite en secuencia todas las ciudades $1, \dots, r$. Una elección posible es entonces una permutación x_1, \dots, x_r de $1, \dots, r$, con la interpretación de que de 0, el agente va a la ciudad x_1 , luego a x_2 , y así sucesivamente. Si suponemos que el agente gana una recompensa no negativa $v(i, j)$ siempre que va directamente de la ciudad i a la ciudad j , entonces los ingresos con la opción $\mathbf{x} = (x_1, \dots, x_r)$ es

$$V(\mathbf{x}) = \sum_{i=1}^r v(x_{i-1}, x_i)$$

Si dos permutaciones son cercanas cuando una se obtiene mediante el intercambio de dos coordenadas de otra, podemos utilizar el temple simulado para aproximar el mejor camino. A saber, comenzamos con cualquier permutación \mathbf{x} y hacemos $X_0 = \mathbf{x}$. Ahora, una vez determinado el n -ésimo estado (es decir, una permutación), $n \geq 0$, generamos una permutación cercana en forma aleatoria [eligiendo I, J como alguno de los $\binom{r}{2}$ valores $i \neq j$, $i, j = 1, \dots, r$, todos con la misma probabilidad, y luego intercambiando los valores de los elementos I -ésimo y J -ésimo de X_n]. Sea \mathbf{y} la permutación generada de esta forma. Entonces, si $V(\mathbf{y}) \geq V(X_n)$, hacemos $X_{n+1} = \mathbf{y}$. En caso contrario, $X_{n+1} = \mathbf{y}$ con probabilidad $(1 + n)^{V(\mathbf{y}) - V(X_n)}$, o es igual a X_n en caso contrario [observe que estamos utilizando $\lambda_n = \log(1 + n)$]. ■

10.5 El algoritmo de muestreo con remuestreo de importancia

El algoritmo de muestreo con remuestreo de importancia, o SIR (*sampling importance resampling*), es un método para generar un vector aleatorio \mathbf{X} cuya función de masa

$$f(\mathbf{x}) = C_1 f_0(\mathbf{x})$$

está dada salvo una constante multiplicativa, simulando una cadena de Markov cuyas probabilidades límites están dadas por una función de masa

$$g(\mathbf{x}) = C_2 g_0(\mathbf{x})$$

que también está dada salvo una constante multiplicativa. Es similar a la técnica de aceptación y rechazo, en la que uno primero genera el valor de un vector aleatorio \mathbf{Y} con densidad g y luego, si $\mathbf{Y} = \mathbf{y}$, acepta este valor con probabilidad $f(\mathbf{y})/cg(\mathbf{y})$, donde c es una constante elegida de modo que $f(\mathbf{x})/cg(\mathbf{x}) \leq 1$, para toda \mathbf{x} . Si el valor no es aceptado, entonces el proceso comienza desde el principio, y el valor \mathbf{X} aceptado en cierto momento tiene densidad f . Sin embargo, como f y g no están especificadas en su totalidad, este método no está disponible.

El método SIR comienza con la generación de m estados sucesivos de una cadena de Markov cuya función de masa de probabilidad límite es g . Denotamos estos valores de estado como $\mathbf{y}_1, \dots, \mathbf{y}_m$. Definimos los "pesos" w_i , $i = 1, \dots, m$, como

$$w_i = \frac{f_0(\mathbf{y}_i)}{g_0(\mathbf{y}_i)}$$

y generamos un vector aleatorio \mathbf{X} tal que

$$P\{\mathbf{X} = \mathbf{y}_j\} = \frac{w_j}{\sum_{i=1}^m w_i}, \quad j = 1, \dots, m$$

Mostraremos que, cuando m es grande, el vector aleatorio \mathbf{X} tiene una función de masa aproximadamente igual a f .

Proposición La distribución del vector \mathbf{X} obtenido mediante el método SIR converge en f cuando $m \rightarrow \infty$.

Demostración Sean \mathbf{Y}_i , $i = 1, \dots, m$, los m vectores aleatorios generados por la cadena de Markov cuya función de masa límite es g , y sean $W_i = f_0(\mathbf{Y}_i)/g_0(\mathbf{Y}_i)$ sus pesos. Para un conjunto fijo de vectores \mathcal{A} , sea $I_i = 1$ si $\mathbf{Y}_i \in \mathcal{A}$ e $I_i = 0$ en caso contrario. Entonces

$$P\{\mathbf{X} \in \mathcal{A} | \mathbf{Y}_i, i = 1, \dots, m\} = \frac{\sum_{i=1}^m I_i W_i}{\sum_{i=1}^m W_i} \quad (10.3)$$

Ahora, por el resultado de la ecuación (10.2) para cadenas de Markov, vemos que cuando $m \rightarrow \infty$,

$$\sum_{i=1}^m I_i W_i / m \rightarrow E_g[IW] = E_g[IW|I=1]P_g\{I=1\} = E_g[W|Y \in \mathcal{A}]P_g\{Y \in \mathcal{A}\}$$

y

$$\sum_{i=1}^m W_i / m \rightarrow E_g[W] = E_g[f_o(Y)/g_o(Y)] = \int \frac{f_o(y)}{g_o(y)} g(y) dy = C_2/C_1$$

Por lo tanto, al dividir el numerador y el denominador de (10.3) entre m tenemos que

$$P\{X \in \mathcal{A} | Y_i, i = 1, \dots, m\} \rightarrow \frac{C_1}{C_2} E_g[W|Y \in \mathcal{A}]P_g\{Y \in \mathcal{A}\}$$

Pero

$$\begin{aligned} \frac{C_1}{C_2} E_g[W|Y \in \mathcal{A}]P_g\{Y \in \mathcal{A}\} &= \frac{C_1}{C_2} E_g\left[\frac{f_o(Y)}{g_o(Y)} | Y \in \mathcal{A}\right]P_g\{Y \in \mathcal{A}\} \\ &= \int_{y \in \mathcal{A}} \frac{f(y)}{g(y)} g(y) dy \\ &= \int_{y \in \mathcal{A}} f(y) dy \end{aligned}$$

Por lo tanto, cuando $m \rightarrow \infty$,

$$P\{X \in \mathcal{A} | Y_i, i = 1, \dots, m\} \rightarrow \int_{y \in \mathcal{A}} f(y) dy$$

lo cual implica, por un resultado matemático conocido como el teorema de convergencia dominada de Lebesgue, que

$$P\{X \in \mathcal{A}\} = E[P\{X \in \mathcal{A} | Y_i, i = 1, \dots, m\}] \rightarrow \int_{y \in \mathcal{A}} f(y) dy$$

lo cual demuestra el resultado. ■

El algoritmo de muestreo con remuestreo de importancia para generar en forma aproximada un vector aleatorio con función de masa f comienza generando variables aleatorias con una función de masa conjunta diferente (como en el *muestreo de importancia*) y luego *obtiene una nueva muestra (remuestrea)* de este conjunto de valores generados, para obtener el vector aleatorio.

Ahora, suponga que queremos estimar $E_f[h(X)]$ para cierta función h . Esto se logra generando primero un gran número de estados sucesivos de una cadena de Markov cuyas probabilidades límites están dadas por g . Si estos estados son y_1, \dots, y_m ,

entonces parecería natural elegir k vectores X_1, \dots, X_k con la distribución de probabilidad

$$P\{X = y_j\} = \frac{w_j}{\sum_{i=1}^m w_i}, \quad j = 1, \dots, m$$

donde k/m es pequeño y $w_i = f_o(y_i)/g_o(y_i)$; luego se emplea $\sum_{i=1}^k h(X_i)/k$ como estimador. Sin embargo, un método mejor consiste no en basar el estimador en un conjunto muestra de k valores, sino en utilizar todo el conjunto de m valores generados y_1, \dots, y_m . Ahora mostraremos que

$$\frac{1}{\sum_{i=1}^m w_i} \sum_{j=1}^m w_j h(y_j)$$

es un mejor estimador de $E_f[h(X)]$ que $\sum_{i=1}^k h(X_i)/k$. Para mostrar esto, observe que

$$E[h(X_i) | y_1, \dots, y_m] = \frac{1}{\sum_{i=1}^m w_i} \sum_{j=1}^m w_j h(y_j)$$

y entonces

$$E\left[\frac{1}{k} \sum_{i=1}^k h(X_i) | y_1, \dots, y_m\right] = \frac{1}{\sum_{i=1}^m w_i} \sum_{j=1}^m w_j h(y_j)$$

lo cual muestra que $\sum_{i=1}^m h(y_i) w_i / \sum_{i=1}^m w_i$ tiene la misma media y menor varianza que $\sum_{i=1}^k h(X_i)/k$.

El uso de datos generados a partir de una distribución para obtener información acerca de otra distribución es de particular utilidad en la estadística bayesiana.

Ejemplo 101 Suponga que X es un vector aleatorio cuya distribución de probabilidad está dada, salvo un vector de parámetros desconocidos θ . Por ejemplo, X podría ser una sucesión de variables aleatorias normales, independientes e idénticamente distribuidas y $\theta = (\theta_1, \theta_2)$, donde θ_1 es la media y θ_2 es la varianza de estas variables aleatorias. Sea $f(x | \theta)$ la densidad de X dado θ . Mientras en la estadística clásica uno supone que θ es un vector de constantes desconocidas, en la estadística bayesiana suponemos que además es aleatorio y que tiene una función de densidad de probabilidad dada $p(\theta)$, conocida como la densidad a priori.

Si se observa que X es igual a x , entonces la densidad condicional, conocida como densidad posterior, de θ está dada por

$$p(\theta | x) = \frac{f(x | \theta)p(\theta)}{\int f(x | \theta)p(\theta)d(\theta)}$$

Sin embargo, en muchos casos es difícil calcular $f(x|\theta)p(\theta)d(\theta)$, de modo que la fórmula anterior no se puede utilizar de manera directa para estudiar la distribución posterior.

Un método para estudiar las propiedades de la distribución posterior consiste en generar primero vectores aleatorios θ a partir de la densidad a priori p y luego utilizar los datos resultantes para obtener información acerca de la densidad posterior $p(\theta|x)$. Si suponemos que la densidad a priori $p(\theta)$ está completamente especificada y que se puede generar directamente a partir de ella, entonces nos servimos de lo anterior, con

$$f_o(\theta) = f(x|\theta)p(\theta)$$

$$g(\theta) = g_o(\theta) = p(\theta)$$

$$w(\theta) = f(x|\theta)$$

Para comenzar, generamos un gran número m de vectores aleatorios a partir de la densidad a priori $p(\theta)$. Suponga que los valores son $\theta_1, \dots, \theta_m$. Ahora podemos estimar cualquier función de la forma $E[h(\theta)|x]$ mediante el estimador

$$\sum_{j=1}^m \alpha_j h(\theta_j), \quad \text{donde } \alpha_j = \frac{f(x|\theta_j)}{\sum_{i=1}^m f(x|\theta_i)}$$

Por ejemplo, para cualquier conjunto \mathcal{A} utilizaríamos

$$\sum_{j=1}^m \alpha_j I\{\theta_j \in \mathcal{A}\} \quad \text{para estimar } P\{\theta \in \mathcal{A} | x\}$$

donde $I\{\theta_j \in \mathcal{A}\}$ es igual a 1 si $\theta_j \in \mathcal{A}$ y es igual a 0 en caso contrario.

Cuando la dimensión de θ es pequeña, podemos utilizar los datos generados a partir de la distribución a priori, junto con sus pesos, para examinar en forma gráfica la distribución a posteriori. Por ejemplo, si θ tiene dimensión 2, podemos graficar los valores generados a priori $\theta_1, \dots, \theta_m$ sobre una gráfica bidimensional, de manera que se tomen en cuenta los pesos de estos puntos. Por ejemplo, podríamos trazar un círculo sobre cada uno de estos m puntos, donde el área del círculo sobre θ_j sería proporcional a su peso $f(x|\theta_j)$. Otra posibilidad sería que todos los círculos fuesen del mismo tamaño, pero donde el tono del círculo dependa de su peso de manera aditiva lineal. Es decir, por ejemplo, si $m=3$ y $\theta_1 = \theta_2$, $f(x|\theta_3) = 2f(x|\theta_1)$, entonces los colores de los círculos sobre θ_1 y θ_3 serían iguales.

Si la densidad a priori p sólo está dada salvo una constante o si es difícil generar vectores aleatorios a partir de ella, entonces podemos generar una cadena de Markov con p como densidad límite, y continuar como antes. ■

Observación El estimador de $E[h(\theta)|x]$ presentado en el ejemplo 10l podría haberse deducido mediante un argumento del tipo de muestreo de importancia. Como

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{C}$$

donde

$$C = \int f(x|\theta)p(\theta)d\theta = E[f(x|\theta)]$$

tenemos que

$$\begin{aligned} E[h(\theta)|x] &= \int \frac{h(\theta)p(\theta|x)}{p(\theta)} p(\theta)d\theta = E\left[\frac{h(\theta)p(\theta|x)}{p(\theta)}\right] \\ &= \frac{E[f(x|\theta)h(\theta)]}{C} = \frac{E[f(x|\theta)h(\theta)]}{E[f(x|\theta)]} \end{aligned}$$

Pero lo anterior sugiere el estimador

$$\frac{\sum_{j=1}^m f(x|\theta_j)h(\theta_j)}{\sum_{j=1}^m f(x|\theta_j)}$$

que es precisamente el estimador dado en el ejemplo 10l. ■

Ejercicios

1. Sean $\pi_j, j=1, \dots, N$, las probabilidades estacionarias de una cadena de Markov. Muestre que si $P\{X_0=j\} = \pi_j, j=1, \dots, N$, entonces

$$P\{X_n=j\} = \pi_j, \quad \text{para toda } n, j$$

2. Sea Q una matriz simétrica de probabilidad de transición; es decir, $q_{ij} = q_{ji}$ para toda i, j . Considere una cadena de Markov tal que, cuando el estado actual es i , genera el valor de una variable aleatoria X tal que $P\{X=j\} = q_{ij}$, y si $X=j$, entonces se mueve al estado j con probabilidad $b_j/(b_i + b_j)$ o permanece en el estado i en caso contrario, donde $b_j, j=1, \dots, N$ son números positivos dados. Muestre que la cadena de Markov resultante es reversible en el tiempo, con probabilidades límites $\pi_j = Cb_j, j=1, \dots, N$.
3. Sea S el conjunto de todas las matrices A de tamaño $n \times n$ cuyos elementos son 0 o 1. (Así, S tiene 2^{n^2} matrices.) Los elementos a_{ij} y a_{rs} de la matriz A son cercanos si $|r-i| + |s-j| = 1$. (Así, por ejemplo, los elementos cercanos

a $a_{2,2}$ son $a_{1,2}$, $a_{2,1}$, $a_{2,3}$ y $a_{3,2}$.) Sea \mathcal{N} el conjunto de todos los pares de elementos cercanos de A . La "energía Ising" de la matriz A se define como

$$H(A) = - \sum_{\mathcal{N}} a_{i,j} a_{r,s}$$

donde la suma es sobre todos los pares de elementos cercanos. Dé un método para elegir de manera aleatoria tal matriz A de acuerdo con la función de masa de probabilidad

$$P(A) = \frac{\exp\{-\lambda H(A)\}}{\sum_{A \in S} \exp\{-\lambda H(A)\}}, \quad A \in S$$

donde λ es una constante positiva dada.

Sugerencia: Digamos que las matrices A y B son cercanas si $A - B$ sólo tiene un elemento distinto de cero.

4. Considere un sistema de 20 componentes independientes, donde el componente i es funcional con probabilidad $0.5 + i/50$, $i = 1, \dots, 20$. Sea X la cantidad de componentes funcionales. Utilice la simulación para estimar la función de masa de probabilidad condicional $P\{X = i | X \leq 5\}$, $i = 1, 2, 3, 4, 5$.

5. Suponga que las variables aleatorias X y Y toman valores en el intervalo $(0, B)$. Suponga que la densidad conjunta de X dado que $Y = y$ es

$$f(x|y) = C(y)e^{-xy}, \quad 0 < x < B$$

y que la densidad conjunta de Y dado que $X = x$ es

$$f(y|x) = C(x)e^{-xy}, \quad 0 < y < B$$

Dé un método para simular aproximadamente el vector X, Y . Ejecute una simulación para estimar (a) $E[X]$ y (b) $E[XY]$.

6. Dé un método eficiente para generar nueve puntos uniformes en $(0, 1)$, condicionando con el evento de que ningún par de ellos estén a menos de 0.1 entre sí. (Se puede mostrar que si n puntos son independientes y uniformemente distribuidos en $(0, 1)$, entonces la probabilidad de que ningún par de ellos estén a menos de una distancia d entre sí es $[1 - (n-1)d]^n$ para $0 < d < 1/(n-1)$.)

7. En el ejemplo 10d, se puede mostrar que la función de masa límite del número de clientes en los $m+1$ servidores es

$$p(n_1, \dots, n_m, n_{m+1}) = C \prod_{i=1}^{m+1} P_i(n_i), \quad \sum_{i=1}^{m+1} n_i = r$$

donde para cada $i = 1, \dots, m+1$, $P_i(n)$, $n = 0, \dots, r$, es una función de masa de probabilidad. Sea e_k el $m+1$ vector unitario, con un 1 en la k -ésima posición y

cero en el resto. Para un vector $\mathbf{n} = (n_1, \dots, n_{m+1})$, sea

$$q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) = \frac{\mathbf{I}(n_i > 0)}{(m+1) \sum_{j=1}^{m+1} \mathbf{I}(n_j > 0)}$$

En palabras, q es la matriz de probabilidades de transición de una cadena de Markov, que en cada paso elige de manera aleatoria un servidor no vacío y luego envía uno de sus clientes a un servidor elegido al azar. Utilice esta función q para proporcionar el algoritmo de Hastings-Metropolis para generar una cadena de Markov con $p(n_1, \dots, n_m, n_{m+1})$ como función de masa límite.

8. Sean X_i , $i = 1, 2, 3$, exponenciales independientes con media 1. Realice un estudio de simulación para estimar

(a) $E[X_1 + 2X_2 + 3X_3 | X_1 + 2X_2 + 3X_3 > 15]$.

(b) $E[X_1 + 2X_2 + 3X_3 | X_1 + 2X_2 + 3X_3 < 1]$.

9. Sea X_1, \dots, X_{20} una permutación aleatoria de $1, \dots, 20$. Estime $P\{\sum_{j=1}^{20} jX_j > 2500\}$ mediante una técnica de Monte Carlo con una cadena de Markov.

10. Suponga que la densidad conjunta de X, Y, Z está dada por

$$f(x, y, z) = Ce^{-(x+y+z+axy+bxz+cyz)}, \quad x > 0, y > 0, z > 0$$

donde a, b, c son constantes no negativas dadas, y C no depende de x, y, z . Explique la forma de simular el vector X, Y, Z y realice una simulación para estimar $E[XYZ]$ cuando $a = b = c = 1$.

11. Suponga que para las variables aleatorias X, Y, N

$$P\{X = i, y \leq Y \leq y + dy, N = n\} \approx C \binom{n}{i} y^{i+\alpha-1} (1-y)^{n+\beta-1} e^{-\lambda} \frac{\lambda^n}{n!} dy$$

donde $i = 0, \dots, n$, $n = 0, 1, \dots$, $y \geq 0$ y α, β, λ son constantes dadas. Ejecute una simulación para estimar $E[X]$, $E[Y]$ y $E[N]$, cuando $\alpha = 2$, $\beta = 3$ y $\lambda = 4$.

12. Utilice el algoritmo SIR para generar una permutación de $1, 2, \dots, 100$ cuya distribución sea aproximadamente la de una permutación aleatoria X_1, \dots, X_{100} condicionando sobre el evento $\sum_j jX_j > 285,000$.

13. Sean X^1, X^2, \dots, X^n puntos aleatorios en \mathcal{C} , el círculo de radio 1 con centro en el origen. Suponga que para cierta r , $0 < r < 1$, su función de densidad conjunta está dada por

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = K \exp\{-\beta t(r; \mathbf{x}_1, \dots, \mathbf{x}_n)\},$$

$$\mathbf{x}_i \in \mathcal{C}, i = 1, \dots, n$$

donde $t(r; \mathbf{x}_1, \dots, \mathbf{x}_n)$ es el número de los $\binom{n}{2}$ pares de puntos $\mathbf{x}_i, \mathbf{x}_j$, $i \neq j$, que están a menos de una distancia r entre sí, y $0 < \beta < \infty$ (observe que $\beta = \infty$ corresponde al caso donde las X^i están uniformemente distribuidas en el círculo, sujetas a la restricción de que no haya dos puntos a menos de una distancia r entre ellos). Explique por qué puede utilizar el algoritmo SIR para generar de manera aproximada estos puntos aleatorios. Si r y β son ambos grandes, ¿sería éste un algoritmo eficiente?

14. Genere 100 números aleatorios $U_{0,k}$, $k = 1, \dots, 10$, $U_{i,j}$, $i \neq j$, $i, j = 1, \dots, 10$. Ahora considere un problema del agente de ventas viajero en el cual éste parte de la ciudad 0 y debe viajar en orden a cada una de las 10 ciudades de acuerdo con alguna permutación de $1, \dots, 10$. Sea U_{ij} la ganancia obtenida por el agente si viaja directamente de la ciudad i a la ciudad j . Emplee el temple simulado para aproximar la máxima ganancia posible del agente de ventas.

Bibliografía

- Aarts, E. y J. Korst, *Simulated Annealing and Boltzmann Machines*, Nueva York, Wiley, 1989.
- Besag, J., "Towards Bayesian Image Analysis", en *Journal of Applied Statistics*, **16**, 1989, pp. 395-407.
- Besag, J., P. Green, D. Higdon y K. Mengersen, "Bayesian Computation and Stochastic Systems (with Discussion)", en *Statistical Science*, **10**, 1995, pp. 3-67.
- Diaconis, P. y S. Holmes, "Three Examples of Monte-Carlo Markov Chains: At the Interface between Statistical Computing, Computer Science, And Statistical Mechanics", en D. Aldous, P. Diaconis, J. Spencer y J. M. Steele (comps.), *Discrete Probability and Algorithms*, Springer Verlag, 1995, pp. 43-56.
- Gelfand, A. E., S. E. Hills, A. Racine-Poon y A. F. Smith, "Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling", en *Journal of the American Statistical Association*, **85**, 1990, pp. 972-985.
- Gelfand A. E. y A. F. Smith, "Sampling Based Approaches to Calculating Marginal Densities", en *Journal of the American Statistical Association*, **85**, 1990, pp. 398-409.
- Gelman, A. y D. B. Rubin, "Inference from Iterative Simulation (with Discussion)", en *Statistical Science*, **7**, 1992, pp. 457-511.

- Geman, S. y D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", en *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 1984, pp. 721-724.
- Geyer, C. J., "Practical Markov Chain Monte Carlo (with Discussion)", en *Statistical Science*, **7**, 1992, pp. 473-511.
- Gidas, B., "Metropolis-type Monte Carlo Simulation Algorithms and Simulated Annealing", en J. L. Sneel (comp.), *Trends in Contemporary Probability*, Boca Ratón, CRC Press, 1995.
- Hajek, B., "Cooling Schedules for Optimal Annealing", en *Math. of Operations Research*, **13**, 1989, pp. 311-329.
- Hammersley, J. M. y D. C. Handscomb, *Monte Carlo Methods*, Londres, Methuen, 1965.
- Ripley, B., *Stochastic Simulation*, Nueva York, Wiley, 1987.
- Rubin, D. R., "Using the SIR Algorithm to Simulate Posterior Distributions", en J. M. Bernardo, M. H. DeGroot, D. V. Lindley y A. F. M. Smith (comps.), *Bayesian Statistics 3*, Oxford University Press, 1988, pp. 395-402.
- Rubinstein, R. R., *Monte Carlo Optimization, Simulation, and Sensitivity of Queueing Networks*, Nueva York, Wiley, 1986.
- Sinclair, A., *Algorithms for Random Generation and Counting*, Boston, Birkhauser, 1993.
- Smith, A. F. y G. O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with Discussion)", en *Jour. of Royal Statistical Society, Series B*, **55**, 1993, pp. 3-23.

Introducción

En este capítulo presentamos algunos temas adicionales, que son un poco más particulares que los considerados en los capítulos anteriores. En la sección 11.1 explicamos una técnica muy conveniente, llamada método de las variables alias, para la generación de variables aleatorias discretas. En la sección 11.2 definimos y presentamos un método para simular un proceso Poisson bidimensional. En la sección 11.3 vemos una identidad probabilística relacionada con la suma de variables aleatorias Bernoulli y mostraremos que su uso puede conducir a reducciones drásticas de la varianza al estimar probabilidades pequeñas. Por último, en la sección 11.4 consideramos el problema de la estimación eficiente de las probabilidades y esperanzas relacionadas con el tiempo de primera pasada de los procesos de Markov haciendo uso de la variable de riesgo total.

11.1 El método de las variables alias para generar variables aleatorias discretas

En esta sección estudiamos una técnica para generar variables aleatorias discretas muy fácil de implantar, aunque precisa cierto tiempo de configuración.

En lo sucesivo, las cantidades $\mathbf{P}, \mathbf{P}^{(k)}, \mathbf{Q}^{(k)}, k \leq n-1$, representan funciones de masa de probabilidad sobre los enteros $1, 2, \dots, n$; es decir, son n -vectores de números no negativos que suman 1. Además, el vector $\mathbf{P}^{(k)}$ tiene a lo más k componentes no nulas, y cada una de las $\mathbf{Q}^{(k)}$ tiene a lo más 2 componentes no nulas. Mostraremos que cualquier función de masa de probabilidad \mathbf{P} se puede representar como una combinación, con

pesos iguales, de $n-1$ funciones de masa de probabilidad \mathbf{Q} (cada una con a lo más dos componentes no nulas). Es decir, mostraremos que para $\mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(n-1)}$ definidas de manera adecuada, \mathbf{P} se puede expresar en la forma

$$\mathbf{P} = \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbf{Q}^{(k)} \quad (11.1)$$

Como preludio a la presentación del método para obtener esta representación, necesitamos el siguiente lema sencillo, cuya demostración se deja como ejercicio.

Lema Sea $\mathbf{P} = \{P_i, i = 1, \dots, n\}$ una función de masa de probabilidad. Entonces

- (a) Hay una $i, 1 \leq i \leq n$, tal que $P_i < 1/(n-1)$, y
- (b) para esta i existe una $j, j \neq i$, tal que $P_i + P_j \geq 1/(n-1)$.

Antes de presentar la técnica general para obtener la representación (11.1), la ilustraremos mediante un ejemplo.

Ejemplo 11a Consideremos la distribución de tres puntos \mathbf{P} , con $P_1 = \frac{7}{16}, P_2 = \frac{1}{2}, P_3 = \frac{1}{16}$. Primero elegimos i y j que satisfagan las condiciones del lema anterior. Como $P_3 < \frac{1}{2}$ y $P_3 + P_2 > \frac{1}{2}$, podemos trabajar con $i = 3$ y $j = 2$. Ahora definimos una función de masa de dos puntos $\mathbf{Q}^{(1)}$, colocando todo su peso en 3 y 2 y tal que \mathbf{P} se pueda expresar como una combinación con pesos iguales de $\mathbf{Q}^{(1)}$ y de una segunda función de masa de dos puntos $\mathbf{Q}^{(2)}$. Además, toda la masa del punto 3 está contenida en $\mathbf{Q}^{(1)}$. Como tenemos que

$$P_j = \frac{1}{2} (Q_j^{(1)} + Q_j^{(2)}), \quad j = 1, 2, 3 \quad (11.2)$$

y, por lo anterior, $Q_3^{(2)}$ se supone que es igual a cero, debemos considerar

$$Q_3^{(1)} = 2P_3 = \frac{1}{8}, \quad Q_2^{(1)} = 1 - Q_3^{(1)} = \frac{7}{8}, \quad Q_1^{(1)} = 0$$

Para satisfacer (10.2), debemos tener

$$Q_3^{(2)} = 0, \quad Q_2^{(2)} = 2P_2 - \frac{7}{8} = \frac{1}{8}, \quad Q_1^{(2)} = 2P_1 = \frac{7}{8}$$

Por lo tanto, en este caso obtenemos la representación deseada. Suponga ahora que la distribución original era la siguiente función de masa de cuatro puntos:

$$P_1 = \frac{7}{16}, \quad P_2 = \frac{1}{4}, \quad P_3 = \frac{1}{8}, \quad P_4 = \frac{3}{16}$$

Ahora, $P_3 < \frac{1}{3}$ y $P_3 + P_1 > \frac{1}{3}$. Por lo tanto, nuestra función de masa inicial de dos puntos ($\mathbf{Q}^{(1)}$) se concentra en los puntos 3 y 1 (sin dar peso a 2 y a 4). Como la representación

final da un peso $\frac{1}{3}$ a $Q^{(1)}$ y además las otras $Q^{(j)}$, $j = 2, 3$, no dan masa alguna al valor 3, debemos tener

$$\frac{1}{3} Q_3^{(1)} = P_3 = \frac{1}{8}$$

Por lo tanto,

$$Q_3^{(1)} = \frac{3}{8}, \quad Q_1^{(1)} = 1 - \frac{3}{8} = \frac{5}{8}.$$

Además, podemos escribir

$$P = \frac{1}{3} Q^{(1)} + \frac{2}{3} P^{(3)}$$

donde $P^{(3)}$, para satisfacer lo anterior, debe ser el vector

$$P_1^{(3)} = \frac{3}{2} \left(P_1 - \frac{1}{3} Q_1^{(1)} \right) = \frac{11}{32}$$

$$P_2^{(3)} = \frac{3}{2} P_2 = \frac{3}{8}$$

$$P_3^{(3)} = 0$$

$$P_4^{(3)} = \frac{3}{2} P_4 = \frac{9}{32}$$

Observe que $P^{(3)}$ no da masa alguna al valor 3. Ahora podemos expresar la función de masa $P^{(3)}$ como una combinación con los mismos pesos de las funciones de masa de dos puntos $Q^{(2)}$ y $Q^{(3)}$, para terminar con

$$\begin{aligned} P &= \frac{1}{3} Q^{(1)} + \frac{2}{3} \left(\frac{1}{2} Q^{(2)} + \frac{1}{2} Q^{(3)} \right) \\ &= \frac{1}{3} (Q^{(1)} + Q^{(2)} + Q^{(3)}) \end{aligned}$$

(dejamos como ejercicio al lector que complete los detalles). ■

El ejemplo anterior bosqueja el siguiente procedimiento general para escribir la función de masa de n puntos P en la forma (11.1), donde cada una de las $Q^{(j)}$ son funciones de masa con toda su masa concentrada en a lo más dos puntos. Para comenzar, elegimos i y j que satisfagan las condiciones del lema. Luego definimos la función de masa $Q^{(1)}$ concentrada en los puntos i y j y la cual contiene toda la masa del

punto i , observando que en la representación (10.1), $Q_i^{(k)} = 0$ para $k = 2, \dots, n-1$, lo cual implica que

$$Q_i^{(1)} = (n-1)P_i \quad \text{y así} \quad Q_j^{(1)} = 1 - (n-1)P_i$$

Al escribir

$$P = \frac{1}{n-1} Q^{(1)} + \frac{n-2}{n-1} P^{(n-1)} \quad (11.3)$$

donde $P^{(n-1)}$ representa la masa restante, vemos que

$$P_i^{(n-1)} = 0$$

$$P_j^{(n-1)} = \frac{n-1}{n-2} \left(P_j - \frac{1}{n-1} Q_j^{(1)} \right) = \frac{n-1}{n-2} \left(P_j + P_i - \frac{1}{n-1} \right)$$

$$P_k^{(n-1)} = \frac{n-1}{n-2} P_k, \quad k \neq i \text{ o } j$$

Podemos verificar fácilmente que lo anterior es realmente una función de masa de probabilidad; por ejemplo, la no negatividad de $P_j^{(n-1)}$ es consecuencia del hecho de que j fue elegido de modo que $P_i + P_j \geq 1/(n-1)$.

Ahora podemos repetir el procedimiento anterior con la función de masa de probabilidad de $(n-1)$ puntos $P^{(n-1)}$ para obtener

$$P^{(n-1)} = \frac{1}{n-2} Q^{(2)} + \frac{n-3}{n-2} P^{(n-2)}$$

y así, de (11.3) tenemos

$$P = \frac{1}{n-1} Q^{(1)} + \frac{1}{n-1} Q^{(2)} + \frac{n-3}{n-1} P^{(n-2)}$$

Ahora repetimos el procedimiento para $P^{(n-2)}$ y así sucesivamente, hasta obtener por último

$$P = \frac{1}{n-1} (Q^{(1)} + \dots + Q^{(n-1)})$$

De esta forma podemos representar P como una combinación con los mismos pesos de $n-1$ funciones de masa de dos puntos. Ahora podemos simular con facilidad a partir de P , generando primero un entero aleatorio N que puede tomar uno de los valores $1, 2, \dots, n-1$, todos con la misma probabilidad. Si el valor resultante N es tal que $Q^{(N)}$ coloca un peso positivo sólo en los puntos i_N y j_N , podemos hacer a X igual a i_N si un segundo número aleatorio es menor que $Q_{i_N}^{(N)}$ e igual a j_N en caso contrario. La

variable aleatoria X tendrá función de masa de probabilidad P . Es decir, tenemos el siguiente procedimiento para simular a partir de P .

PASO 1: Generar U_1 y hacer $N = 1 + \text{Ent}[(n-1)U_1]$.

PASO 2: Generar U_2 y hacer

$$X = \begin{cases} i_N & \text{si } U_2 < Q_{i_N}^{(N)} \\ j_N & \text{en caso contrario} \end{cases}$$

Observaciones

1. Lo anterior se llama método de las variables alias debido a que una nueva numeración de las Q siempre permite ordenar las cosas de modo que para cada k , $Q_k^{(k)} > 0$ (es decir, podemos arreglar las cosas de modo que la k -ésima función de masa de dos puntos asigne un peso positivo al valor k). Por lo tanto, el procedimiento requiere la simulación de N , para que tome alguno de los valores $1, 2, \dots, n-1$, cada uno con la misma probabilidad, y luego acepta a k como el valor de X si $N = k$, o acepta como valor de X el "alias" de k (a saber, el otro valor al cual $Q^{(k)}$ le da peso positivo).
2. En realidad, no es necesario generar un nuevo número aleatorio en el paso 2. Como $N-1$ es la parte entera de $(n-1)U_1$, tenemos que la diferencia $(n-1)U_1 - (N-1)$ es independiente de N_1 y está uniformemente distribuida en $(0, 1)$. Por lo tanto, en vez de generar un nuevo número aleatorio U_2 en el paso 2, podemos utilizar $(n-1)U_1 - (N-1)$. ■

11.2 Simulación de un proceso Poisson bidimensional

Un proceso formado por puntos aleatorios en el plano es un proceso Poisson bidimensional con razón λ , $\lambda > 0$, si

1. El número de puntos que aparecen en cualquier región dada de área A se distribuye como Poisson con media λA .
2. El número de puntos que aparecen en regiones ajenas son independientes.

Para un punto fijo dado 0 del plano, mostraremos cómo simular los puntos que aparecen en una región circular de radio r con centro en 0 , de acuerdo con un proceso Poisson bidimensional con razón λ .

Sea $C(a)$ el círculo de radio a con centro en 0 , y observemos que, por la condición 1, el número de puntos en $C(a)$ se distribuye como Poisson con media $\lambda \pi a^2$. Sea

R_i , $i \geq 1$, la distancia del origen 0 a su i -ésimo punto más cercano (véase la figura 11.1). Entonces

$$\begin{aligned} P\{\pi R_1^2 > x\} &= P\{R_1 > \sqrt{x/\pi}\} \\ &= P\left\{\text{no haya puntos en } C\left(\sqrt{\frac{x}{\pi}}\right)\right\} \\ &= e^{-\lambda x} \end{aligned}$$

donde la última igualdad utiliza el hecho de que el área de $C(\sqrt{x/\pi})$ es x . Además, si $C(b) - C(a)$ denota la región entre $C(b)$ y $C(a)$, $a < b$, tenemos

$$\begin{aligned} P\{\pi R_2^2 - \pi R_1^2 > x | R_1 = a\} &= P\left\{R_2 > \sqrt{\frac{x + \pi R_1^2}{\pi}} | R_1 = a\right\} \\ &= P\left\{\text{no haya puntos en } C\left(\sqrt{\frac{x + \pi a^2}{\pi}}\right) - C(a) | R_1 = a\right\} \\ &= P\left\{\text{no haya puntos en } C\left(\sqrt{\frac{x + \pi a^2}{\pi}}\right) - C(a)\right\} \quad \text{por condición 2} \\ &= e^{-\lambda x} \end{aligned}$$

De hecho, podemos repetir este argumento en forma continua para obtener la siguiente proposición.

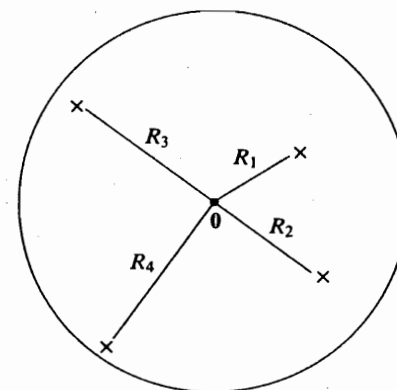


Figura 11.1

Proposición Con $R_0 = 0$, $\pi R_i^2 - \pi R_{i-1}^2$, $i \geq 1$, son variables aleatorias exponenciales independientes, cada una con razón λ .

En otras palabras, la cantidad de área que debe recorrerse para encontrar un punto Poisson es exponencial con razón λ . Como por simetría, los ángulos respectivos de los puntos Poisson son independientes y distribuidos uniformemente en $(0, 2\pi)$, tenemos el siguiente algoritmo para simular el proceso Poisson sobre una región circular de radio r con centro en $\mathbf{0}$.

PASO 1: Generar exponenciales independientes con razón λ , X_1, X_2, \dots , deteniéndolos en

$$N = \text{Mín}\{n: X_1 + \dots + X_n > \pi r^2\}$$

PASO 2: Si $N = 1$, no hay puntos en $C(r)$ y el algoritmo termina. En caso contrario, para $i = 1, \dots, N-1$, hacer

$$R_i = \sqrt{\frac{X_1 + \dots + X_i}{\pi}}$$

(es decir, $\pi R_i^2 = X_1 + \dots + X_i$).

PASO 3: Generar números aleatorios U_1, \dots, U_{N-1} .

PASO 4: Las coordenadas polares de los $N-1$ puntos Poisson son

$$(R_i, 2\pi U_i), \quad i = 1, \dots, N-1$$

Podemos considerar el algoritmo anterior como el barrido de un círculo con centro en $\mathbf{0}$, con un radio que varía continuamente de 0 hasta r . Los radios donde se encuentran los puntos se simulan utilizando el resultado de que el área que debe recorrerse hasta encontrar otro punto siempre se distribuye en forma exponencial con razón λ . Esta técnica de barrido también sirve para simular el proceso en regiones no circulares. Por ejemplo, consideremos una función no negativa $f(x)$ y suponga que estamos interesados en simular el proceso Poisson en la región comprendida entre el eje x y la función f (véase la figura 11.2), donde x va de 0 hasta T . Para esto, podemos comenzar en la orilla izquierda y barrir la región en forma vertical hasta la derecha, analizando las diversas áreas sucesivas. Específicamente, si $X_1 < X_2 < \dots$ denotan las proyecciones sucesivas del proceso Poisson sobre el eje x , tenemos, exactamente de la misma manera que antes, que (con $X_0 = 0$)

$$\int_{X_{i-1}}^{X_i} f(x) dx, \quad i = 1, \dots, \text{son exponenciales independientes con razón } \lambda.$$

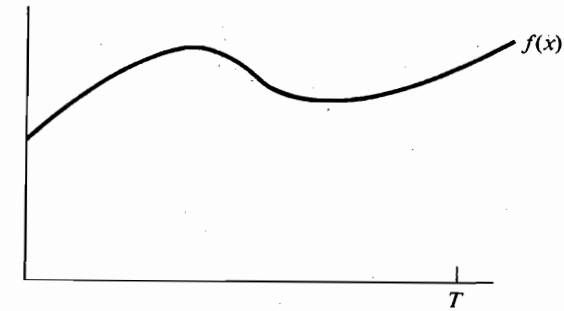


Figura 11.2

Por lo tanto, para simular lo anterior generamos variables aleatorias exponenciales independientes con razón λ , W_1, W_2, \dots , y nos detenemos en

$$N = \text{Mín}\left\{n: W_1 + \dots + W_n > \int_0^T f(x) dx\right\}$$

Ahora debemos determinar X_1, \dots, X_{N-1} mediante las ecuaciones

$$\int_0^{X_1} f(x) dx = W_1$$

$$\int_{X_1}^{X_2} f(x) dx = W_2$$

\vdots

$$\int_{X_{N-2}}^{X_{N-1}} f(x) dx = W_{N-1}$$

Como es claro que la proyección sobre el eje y del punto cuya coordenada x es X_i se distribuye de manera uniforme sobre $(0, f(X_i))$, si generamos números aleatorios U_1, \dots, U_{N-1} , entonces los puntos Poisson simulados son, en coordenadas rectangulares, $(X_i, U_i f(X_i))$, $i = 1, \dots, N-1$.

El procedimiento anterior es más útil cuando f es lo suficientemente regular que es posible despejar de manera eficiente los valores de X_i en las ecuaciones anteriores. Por ejemplo, si $f(x) = c$ (de modo que la región es un rectángulo), podemos expresar a X_i como

$$X_i = \frac{W_1 + \dots + W_i}{c}$$

y los puntos Poisson son

$$(X_i, cU_i), \quad i = 1, \dots, N-1$$

11.3 Aplicaciones de una identidad para la suma de variables aleatorias Bernoulli a la simulación

Sean X_1, \dots, X_m variables aleatorias Bernoulli tales que

$$P\{X_i = 1\} = \lambda_i = 1 - P\{X_i = 0\}.$$

Además, sean $S = \sum_{i=1}^m X_i$, y $\lambda = E[S] = \sum_{i=1}^m \lambda_i$. Sea R una variable aleatoria arbitraria, y suponga que I , independiente de R, X_1, \dots, X_m , es tal que

$$P\{I = i\} = 1/m, \quad i = 1, \dots, m$$

Es decir, I es una variable aleatoria discreta uniforme en $1, \dots, m$ que es independiente de las demás variables aleatorias.

La siguiente identidad es la clave para los resultados de esta sección.

Proposición

- (a) $P\{I = i | X_i = 1\} = \lambda_i / \lambda$
- (b) $E[SR] = \lambda E[R | X_i = 1]$
- (c) $P\{S > 0\} = \lambda E[1/S | X_i = 1]$

Demostración Para demostrar la parte (a), observe que

$$P\{I = i | X_i = 1\} = \frac{P\{X_i = 1 | I = i\}P\{I = i\}}{\sum_i P\{X_i = 1 | I = i\}P\{I = i\}}$$

Ahora,

$$\begin{aligned} P\{X_i = 1 | I = i\} &= P\{X_i = 1\} \quad \text{por independencia} \\ &= \lambda_i \end{aligned}$$

lo cual concluye la demostración de la parte (a). Para demostrar la parte (b), razonemos como sigue:

$$\begin{aligned} E[SR] &= E\left[R \sum_i X_i\right] \\ &= \sum_i E[RX_i] \\ &= \sum_i \{E[RX_i | X_i = 1]\lambda_i + E[RX_i | X_i = 0](1 - \lambda_i)\} \\ &= \sum_i \lambda_i E[R | X_i = 1] \end{aligned} \quad (11.4)$$

Además,

$$\begin{aligned} E[R | X_i = 1] &= \sum_i E[R | X_i = 1, I = i]P\{I = i | X_i = 1\} \\ &= \sum_i E[R | X_i = 1, I = i]\lambda_i / \lambda \quad \text{por la parte (a)} \\ &= \sum_i E[R | X_i = 1]\lambda_i / \lambda \end{aligned} \quad (11.5)$$

La combinación de las ecuaciones (11.4) y (11.5) demuestra la parte (b).

Para demostrar (c), definimos R como 0 si $S = 0$ y como $1/S$ si $S > 0$. Entonces,

$$E[SR] = P\{S > 0\} \quad \text{y} \quad E[R | X_i = 1] = E\left[\frac{1}{S} | X_i = 1\right]$$

de modo que (c) es consecuencia directa de (b). ■

Ahora utilizaremos la proposición anterior para estimar (a) la probabilidad de falla de un sistema y (b) la probabilidad de que cierto patrón ocurra dentro de un lapso determinado.

Ejemplo 11b Consideremos el modelo del ejemplo 8b, relativo a un sistema con n componentes independientes, y suponga que queremos estimar la probabilidad de que el sistema falle, cuando esta probabilidad es muy pequeña. Ahora, para cualquier sistema del tipo considerado en el ejemplo 8b, siempre existe una única familia de conjuntos $\{C_1, \dots, C_m\}$, de los que ninguno es subconjunto de otro, tal que el sistema fallará si y sólo si todos los componentes de al menos uno de estos conjuntos ha fallado. Se trata de los *conjuntos de corte mínimo* del sistema.

Sea $Y_j, j = 1, \dots, n$ igual a 1 si el componente j falla e igual a 0 en caso contrario, y sea $q_j = P\{Y_j = 1\}$ la probabilidad de que el componente j falle. Para $i = 1, \dots, m$, sea

$$X_i = \prod_{j \in C_i} Y_j$$

Es decir, X_i es el indicador del evento de que todos los componentes en C_i han fallado. Si $S = \sum_i X_i$, entonces θ , la probabilidad de que el sistema falle, está dada por

$$\theta = P\{S > 0\}$$

Ahora mostraremos cómo utilizar la proposición para estimar θ de manera eficiente.

En primer lugar, sean $\lambda_i = E[X_i] = \prod_{j \in C_i} q_j$, y $\lambda = \sum_i \lambda_i$. Ahora, simulamos el valor de J , una variable aleatoria que es igual a i con probabilidad λ_i / λ , $i = 1, \dots, m$ [la parte

(a) de la proposición implica que J tiene la misma distribución que la distribución condicional de I , dado que $X_i = 1$. Entonces hacemos Y_i igual a 1 para toda $i \in C$, y simulamos el valor de las demás $Y_i, i \notin C$, haciéndolas iguales a 1 con probabilidad q , e iguales a cero en caso contrario. Sea S^* el número resultante de conjuntos de corte mínimo que tienen todos sus componentes descompuestos, y observe que $S^* \geq 1$. La parte (c) de la proposición implica que λ/S^* es un estimador insesgado de θ . Como $S^* \geq 1$, también se tiene que

$$0 \leq \lambda/S^* \leq \lambda$$

y entonces, cuando λ , el número medio de conjuntos de corte mínimo que están descompuestos, es muy pequeño, el estimador λ/S^* tendrá una varianza muy pequeña.

Por ejemplo, consideremos un sistema tres de cinco que falla si al menos tres de los cinco componentes fallan, y suponga que cada componente puede fallar de manera independiente con probabilidad q . Para este sistema, los conjuntos de corte mínimo serán los $\binom{5}{3} = 10$ subconjuntos de tamaño 3. Como todas las fallas de componentes son las mismas, el valor de I no juega papel alguno. Así, podemos obtener la estimación anterior suponiendo que los componentes 1, 2 y 3 han fallado y generando después el estado de los otros dos. Así, al considerar el número de componentes 4 y 5 que han fallado, y como $\lambda = 10q^3$, se tiene que la distribución del estimador es

$$\begin{aligned} P\{\lambda/S^* = 10q^3\} &= (1 - q)^2 \\ P\{\lambda/S^* = 10q^3/4\} &= 2q(1 - q) \\ P\{\lambda/S^* = q^3\} &= q^2 \end{aligned}$$

Por lo tanto, si $p = 1 - q$,

$$\begin{aligned} \text{Var}(\lambda/S^*) &= E[(\lambda/S^*)^2] - (E[\lambda/S^*])^2 \\ &= 100q^6[p^2 + pq/8 + q^2/100 - (p^2 + pq/2 + q^2/10)^2] \end{aligned}$$

La siguiente tabla proporciona el valor de θ y el cociente de $\text{Var}(I)$ entre la varianza del estimador λ/S^* para varios valores de q , donde $\text{Var}(I) = \theta(1 - \theta)$ es la varianza del estimador de simulación bruto.

q	θ	$\text{Var}(I)/\text{Var}(\lambda/S^*)$
0.001	9.985×10^{-9}	8.896×10^{10}
0.01	9.851×10^{-6}	8.958,905
0.1	0.00856	957.72
0.2	0.05792	62.59
0.3	0.16308	12.29

Así, para q pequeña, $\text{Var}(\lambda/S^*)$ es casi del orden de θ^2 , mientras que $\text{Var}(I) \approx \theta$. ■

Ejemplo 11c Espera de un patrón. Sea $Y_i, i \geq 1$, una sucesión de variables aleatorias discretas, independientes e idénticamente distribuidas, con función de masa de probabilidad $P_j = P\{Y_i = j\}$. Sea s_1, \dots, s_k una sucesión fija de valores posibles de estas variables aleatorias y defina

$$N = \min\{i: i \geq k, Y_{i-j} = s_{k-j}, j = 0, 1, \dots, k-1\}$$

Es decir, N es la primera vez que aparece el patrón s_1, \dots, s_k . Estamos interesados en utilizar la simulación para estimar $\theta = P\{N \leq n\}$, cuando θ es pequeña. Aunque el estimador de simulación usual se obtiene simulando la sucesión de variables aleatorias hasta que aparece el patrón o hasta que ya no sea posible que ocurra hasta el instante n (haciendo que el estimador de esa ejecución sea igual a 1 en el primer caso y 0 en el segundo), mostraremos la forma de aplicar la proposición anterior para obtener un estimador de simulación más eficiente.

Para comenzar, generamos las n variables aleatorias Y_1, \dots, Y_n , y para $i \geq k$ hacemos

$$X_i = 1 \quad \text{si} \quad Y_i = s_k, \quad Y_{i-1} = s_{k-1}, \dots, Y_{i-k+1} = s_1$$

e igual a cero en caso contrario. En otras palabras, X_i es igual a 1 si el patrón ocurre (no necesariamente por primera vez) en el instante i . Sea

$$S = \sum_{i=k}^n X_i$$

el número de veces que ha ocurrido el patrón hasta el instante n y observe que

$$\theta = P\{N \leq n\} = P\{S > 0\}$$

Como para $k \leq i \leq n$

$$\lambda_i = P\{X_i = 1\} = P_{s_1} P_{s_2} \cdots P_{s_k}$$

la proposición implica que

$$\theta = (n - k + 1) \left(\prod_{i=1}^k P_{s_i} \right) E \left[\frac{1}{S} \middle| X_i = 1 \right]$$

donde I , independiente de las Y_j , tome uno de los valores k, \dots, n , cada uno con la misma probabilidad de ocurrir. Así, para estimar θ simulamos primero J , igualmente probable a cualquiera de los valores k, \dots, n , y hacemos

$$Y_J = s_k, \quad Y_{J-1} = s_{k-1}, \dots, Y_{J-k+1} = s_1$$

Luego simulamos los otros $n - k$ valores Y_i de acuerdo con la función de masa P y definimos S^* como el número de veces que ocurre el patrón. El estimador de simulación de θ a partir de esta ejecución es

$$\hat{\theta} = \frac{n - k + 1}{S^*} \prod_{i=1}^k P_{s_i}$$

Para valores pequeños de $(n - k + 1)(\prod_{i=1}^k P_{s_i})$, lo anterior será un estimador muy eficiente de θ . ■

11.4 Estimación de probabilidades y tiempos esperados de primera pasada mediante riesgos aleatorios

Consideremos un proceso que pasa de un estado a otro de la siguiente manera. Si el estado actual es x , entonces, independientemente de la sucesión de estados previos, el siguiente estado se elige de acuerdo con la función de distribución F_x . Tal proceso es un proceso de Markov (si el espacio de estados es discreto, entonces es una cadena de Markov).

Dado un estado inicial 0 y un conjunto fijo de estados A , suponga que estamos interesados en estimar el número esperado de transiciones necesarias para que el proceso entre al conjunto A . Es decir, si X_n es el estado en el instante n , entonces estamos interesados en $E[N]$, donde

$$N = \text{Mín}\{n: X_n \in A\}$$

En muchos casos de interés, A es un “conjunto raro de estados”, en el sentido de que el proceso tiene una tendencia al estado inicial 0 donde las visitas a A ocurren con poca frecuencia. Por lo tanto, en tales casos, $E[N]$ es relativamente grande. Además, como el proceso tiene esta tendencia a regresar a 0, tenemos, por la propiedad de no tener memoria, que N será aproximadamente una distribución exponencial (véase el ejemplo 8a del capítulo 8, para un análisis detallado de este punto). Pero si esto ocurre, tenemos que $\text{Var}(N) \approx (E[N])^2$, lo cual implica que $\sqrt{\text{Var}(N)}$, la desviación estándar de N , también es grande; así, el método bruto de simulación para estimar $E[N]$ necesita un gran número de ejecuciones.

Una poderosa técnica de reducción de la varianza se vale de los llamados riesgos observados o proyecciones predecibles como variables de control. Funciona de la manera siguiente. Considere una ejecución de simulación que produzca la sucesión de estados X_0, \dots, X_N y defina los riesgos observados $\lambda_n, n \geq 0$, como

$$\lambda_n = P\{N = n | X_0, \dots, X_{n-1}\}$$

En palabras, si no se ha entrado a A antes del instante n , entonces λ_n es la probabili-

dad, dado el estado en el instante $n - 1$, de que el siguiente estado esté en el conjunto A . Ahora, sea Y la suma de estos riesgos; es decir,

$$Y = \sum_{n=1}^N \lambda_n$$

Proposición

$$E\left[\sum_{n=1}^N \lambda_n\right] = P\{N < \infty\}.$$

Demostración Sea

$$I_n = \begin{cases} 1 & \text{si } N = n \\ 0 & \text{en caso contrario} \end{cases}$$

y observe que

$$E[I_n | X_0, \dots, X_{n-1}] = P\{N = n | X_0, \dots, X_{n-1}\} = \lambda_n$$

Ahora,

$$\sum_{n=1}^{\infty} I_n = \begin{cases} 1 & \text{si } N < \infty \\ 0 & \text{en caso contrario} \end{cases}$$

y así, al calcular las esperanzas,

$$\begin{aligned} P\{N < \infty\} &= E\left[\sum_{n=1}^{\infty} I_n\right] \\ &= \sum_{n=1}^{\infty} E[I_n] \\ &= \sum_{n=1}^{\infty} E[E[I_n | X_0, \dots, X_{n-1}]] \\ &= E\left[\sum_{n=1}^{\infty} E[I_n | X_0, \dots, X_{n-1}]\right] \\ &= E\left[\sum_{n=1}^{\infty} \lambda_n\right] \\ &= E\left[\sum_{n=1}^N \lambda_n\right] \end{aligned}$$

■

Suponga ahora que la situación es tal que se entra a \mathcal{A} en algún momento con probabilidad 1, de modo que la suma esperada de los riesgos aleatorios tiene media 1. Como las ejecuciones con grandes valores de N surgen con frecuencia de muchas oportunidades fallidas para entrar a A , tenemos que Y es un candidato natural para ser variable de control, pues da una forma de calibrar si un valor dado de N es mucho mayor o mucho menor que lo normal. Por lo tanto, debemos simular hasta que el proceso entre a A y luego utilizar el estimador de salida

$$N + c(Y - 1)$$

El mejor valor de c , dado por

$$c = -\frac{\text{Cov}(N, Y)}{\text{Var}(Y)}$$

se puede aproximar a partir de los datos simulados.

Ejemplo 11d Diagrama de control de suma acumulativa para variables aleatorias exponenciales. Sea X_1, X_2, \dots una sucesión de variables aleatorias exponenciales independientes, cada una con media 1, Sea $S_0 = 0$ y definamos

$$S_n = \text{Máx}\{0, S_{n-1} + X_n - 2\}, \quad n \geq 1$$

Estamos interesados en estimar $E[N]$, donde

$$N = \text{Mín}\{n : S_n > 4\}$$

La sucesión anterior S_n , $n \geq 1$, es un diagrama de control de suma acumulativa. La idea es que las cantidades X_n representan los valores sucesivos de artículos fabricados en cierto proceso. Cuando el proceso está bajo control, se supone que estos valores tienen una distribución exponencial con media 1, mientras que si el proceso está fuera de control, estos valores tienden a ser cada vez mayores. Las cantidades S_n son sumas acumulativas de las cantidades sucesivas $X_n - 2$, las cuales no pueden ser negativas; es decir, el valor de la suma acumulativa es igual a 0 siempre que llegue a ser negativa. Como los valores $X_n - 2$ tienen una media negativa cuando el proceso está bajo control, lo usual es que la suma acumulativa tenga un valor pequeño en este caso, de modo que un valor grande es un indicador de que el proceso puede haberse salido de control. Sin embargo, como podrían ocurrir valores grandes incluso cuando el proceso esté bajo control, estamos interesados en $E[N]$, el tiempo esperado hasta el cual el proceso pueda declararse de manera errónea como fuera de control, siendo que ha permanecido todo el tiempo bajo control.

Como los valores S_n tienden a estar cerca de 0, llegar hasta un valor tan grande como 4 tiende a ser un evento raro con distribución aproximadamente exponencial.

Así, el estimador de simulación bruto tiene una desviación estándar grande. El riesgo aleatorio está dado por

$$\begin{aligned} \lambda_n &= P\{S_{n-1} + X_n - 2 > 4 | S_{n-1}\} \\ &= P\{X_n > 6 - S_{n-1} | S_{n-1}\} \\ &= \exp\{S_{n-1} - 6\} \end{aligned}$$

Por lo tanto, el uso del estimador controlado

$$N + c[Y - 1]$$

donde se recomienda

$$Y = \sum_{n=1}^N \exp\{S_{n-1} - 6\}$$

De hecho, un pequeño estudio de simulación reveló al autor que la varianza del estimador controlado es menor que la del estimador bruto por un factor aproximado de 1/30. Además, se vio que $E[N] \approx 243$. ■

Ejemplo 11e Longitud esperada de un periodo ocupado en un sistema de línea de espera. Considere una línea de espera con un servidor, en la que los tiempos entre las llegadas de los clientes son variables aleatorias independientes con una distribución común F (en otras palabras, el proceso de llegada es un proceso de renovación; véase la sección 8.3 del capítulo 8). Hay un único servidor que tarda un tiempo aleatorio en atender a un cliente. Una llegada que encuentra desocupado al servidor entra de inmediato a servicio, mientras que una que lo encuentra ocupado se une a la fila de espera. Al terminar un servicio, uno de los clientes en espera, si los hay, pasa a servicio. Cada servicio tarda un tiempo aleatorio, independiente del pasado, con distribución G .

El sistema descrito de esta forma alterna entre momentos en los cuales el servidor está ocupado y momentos en los que está desocupado. Suponga que estamos interesados en determinar $E[N]$, el número esperado de clientes atendidos en un periodo arbitrario de ocupación.

Sea W_n la cantidad total de trabajo restante en el sistema; es decir, la suma de los tiempos de servicio de los clientes que esperan, más el tiempo de servicio restante del cliente en servicio, inmediatamente después de la llegada del n -ésimo cliente. Además, sea X_n el número de clientes en el sistema hasta la n -ésima llegada, de modo que $X_1 = 0$ y

$$N = \text{Mín}\{n > 1 : X_n = 0\} - 1$$

Ahora, consideremos el proceso cuyo estado en el instante n es el par W_n, X_n y definamos los riesgos aleatorios como

$$\lambda_n = P\{N = n | W_1, X_1, \dots, W_n, X_n\}$$

Ahora, dado que la carga de trabajo inmediatamente después de la n -ésima llegada es w , tenemos que la siguiente llegada encontrará al sistema vacío (lo cual implica que $N = n$) si el siguiente intervalo entre llegadas es mayor que w . Por lo tanto, para $n \leq N$,

$$\lambda_n = 1 - F(W_n) = \bar{F}(W_n)$$

Así, recomendamos el uso del estimador

$$N + c \left[\sum_{n=1}^N \bar{F}(W_n) - 1 \right]$$

donde el valor de c se puede determinar a partir de la simulación.

Si el sistema de línea de espera tiene más de un servidor, podríamos proceder como antes, excepto que ahora definiríamos W_n como el tiempo hasta el cual el sistema se vacía de todos los clientes presentes en el sistema en el momento inmediato posterior a la llegada del n -ésimo cliente, siempre que no haya llegadas adicionales. Entonces, el estimador se daría como antes.

Si quisiéramos estimar $E[T]$, el tiempo medio de un periodo de ocupación, podríamos estimar $E[N]$ y utilizar la identidad

$$E[T] = E[N]\mu_F$$

donde μ_F es el tiempo medio entre llegadas, o podríamos emplear de manera directa el estimador T (de nuevo, con lo anterior como control). Es decir, podríamos utilizar

$$T + c \left[\sum_{n=1}^N \bar{F}(W_n) - 1 \right]$$

donde c es igual al negativo de la covarianza entre T y el control, dividido entre la varianza del control. Aunque no es inmediatamente claro cuál de estos métodos es mejor (estimar $E[N]$ y luego utilizar $E[N]\mu_F$ o estimar de manera directa $E[T]$), el autor espera que, en general, el primer método sea mejor. ■

Como

$$E \left[\sum_{n=1}^N \lambda_n \right] = P\{N < \infty\}$$

también podemos utilizar la suma de los riesgos como un estimador de $P\{N < \infty\}$ en

casos donde esta probabilidad sea menor que 1. Por ejemplo, si N representa la primera vez que el proceso entra a un estado en \mathcal{A} , entonces

$$P\{N \leq m\} = E \left[\sum_{n=1}^{\min(N,m)} \lambda_n \right]$$

de modo que podemos estimar $P\{N \leq m\}$ mediante $\sum_{n=1}^{\min(N,m)} \lambda_n$. Si existe un conjunto de estados \mathcal{B} tal que si el proceso entra a \mathcal{B} , entonces es imposible que entre a \mathcal{A} de ahí en adelante, entonces

$$P\{\text{entrar en algún momento a } \mathcal{A}\} = E \left[\sum_{n=1}^{\min(N_{\mathcal{A}}, N_{\mathcal{B}})} \lambda_n \right]$$

donde $N_{\mathcal{A}}$ y $N_{\mathcal{B}}$ son, respectivamente, los primeros instantes en que el proceso entra a los estados \mathcal{A} y \mathcal{B} .

Ejemplo 11f Otra forma de utilizar la simulación para estimar $E[N]$ en el ejemplo 11d consiste en definir un ciclo que ocurre cuando S_n es mayor que 4 o cuando regresa a 0. Sea C el tiempo de un ciclo. Entonces es fácil ver que

$$E[N] = \frac{E[C]}{p}$$

donde p es la probabilidad de que la caminata aleatoria sea mayor que 4 antes de que sea negativa. Como p tiende a ser pequeña, parecería que podríamos estimarla de manera eficiente utilizando como estimador la suma de los riesgos de N durante un ciclo. Es decir, podríamos utilizar el estimador

$$\hat{p} = \sum_{n=1}^C \exp\{S_{n-1} - 6\}$$

Además, como es probable que $E[C]$ sea relativamente pequeña, se puede estimar utilizando C junto con la suma de los riesgos como control. Es decir, podemos estimar $\mu_C = E[C]$ mediante un estimador de la forma

$$\hat{\mu}_C = C + c \left[\sum_{n=1}^C (\exp\{S_{n-1} - 6\} + 1 - \exp\{S_{n-1} - 2\}) - 1 \right]$$

Luego nos valemos del cociente $\hat{p}/\hat{\mu}_C$ para estimar $E[N]$. ■

El estimador de riesgo total de $P\{N < \infty\}$ mejora con las ideas de muestreo estratificado. Si $H = \sum_{n=1}^N \lambda_n$, podemos escribir

$$H = \lambda_1 + H_R$$

donde H_R es el riesgo total restante después de la transición inicial. Al condicionar sobre si $X_1 \in \mathcal{A}$ obtenemos

$$E[H] = \lambda_1 + (1 - \lambda_1)E[H_R | X_1 \notin \mathcal{A}]$$

Por lo tanto, si H_1 tiene la distribución condicional de H_R dado que $X_1 \notin \mathcal{A}$, entonces $\lambda_1 + (1 - \lambda_1)H_1$ tiene la misma media y menor varianza que H . La variable H_1 se puede generar simulando la transición inicial del proceso, y si conduce a un estado en \mathcal{A} , entonces este paso se repite hasta que se obtiene un estado que no está en \mathcal{A} . La cantidad H_1 es entonces la suma de los riesgos restantes a partir de este punto. Sin embargo, en vez de utilizar de manera directa a H_1 , podemos mejorar el procedimiento repitiendo el argumento. La repetición continua nos conduce al siguiente algoritmo para generar un estimador insesgado de $P\{N < \infty\}$ cuya varianza es menor o igual a la de H .

1. $n = 0, X = 0$.
2. $X_n = X$.
3. Dar como salida X_n .
4. Generar X con la distribución F_{X_n} .
5. Si $X \in \mathcal{A}$, ir al paso 4.
6. $n = n + 1$.
7. Ir al paso 2.

El algoritmo debe detenerse si se vuelve imposible entrar a \mathcal{A} en el marco de tiempo de interés o si se alcanza un estado que vaya a \mathcal{A} con probabilidad 1. Si $X_n, n \geq 0$, son los valores sucesivos producidos por el algoritmo anterior, entonces la estimación de $p = P\{N < \infty\}$ es

$$\hat{p} = \lambda_1 + \sum_{i=2} \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j)$$

donde

$$\lambda_i = F_{X_{i-1}}(\mathcal{A}) = \int_{x \in \mathcal{A}} dF_{X_{i-1}}(x)$$

Ejercicios

1. Configure el método de las variables alias para generar una binomial con parámetros (5, 0.4).
2. Explique cómo numerar los $Q^{(k)}$ en el método de las variables alias de modo que k sea uno de los dos puntos a los cuales $Q^{(k)}$ da un peso.
3. Complete los detalles del ejemplo 11a.
4. Escriba un programa para generar los puntos de un proceso Poisson bidimensional dentro de un círculo de radio R , y ejecute el programa para $\lambda = 1$ y $R = 5$. Grafique los puntos obtenidos.
5. Utilice una simulación para estimar la probabilidad de que la estructura de puente dada en la figura 8.1 falle si los componentes 1, 2 y 3, de manera independiente, fallan con probabilidad 0.05, y los componentes 4 y 5 con probabilidad 0.01. Además, compare la varianza de su estimador con la del estimador de simulación bruto.
6. Realice una simulación para estimar la probabilidad de que se obtengan 10 caras consecutivas durante los primeros 100 lanzamientos de una moneda. Además, compare la varianza de su estimador con la del estimador de simulación bruto.
7. Utilice una simulación para estimar la probabilidad de que ocurra el patrón cara, cruz, cruz, cara, cruz, dentro de los primeros 20 lanzamientos de una moneda, donde la probabilidad de que caiga cara en el lanzamiento i es $(i + 10)/40$, $i = 1, \dots, 20$. Suponga la independencia.
8. Explique la forma en que se puede reducir la varianza del estimador de simulación del ejemplo 11b mediante variables antitéticas.
9. En el ejercicio 5, determine la reducción de varianza adicional, utilizando variables antitéticas.
10. Sean X_i variables aleatorias Bernoulli con medias λ_i , $i = 1, \dots, m$ y sea $S = \sum_i a_i X_i$, donde las a_i son constantes positivas. Sea R una variable aleatoria arbitraria, y sea I , independiente de las demás variables, con función de masa $P\{I = i\} = a_i / \sum_i a_i$, $i = 1, \dots, m$. Sea $\lambda = \sum_i a_i \lambda_i$.
 - (a) Determine $P\{I = i | X_i = 1\}$.
 - (b) Muestre que $E[SR] = \lambda E[R | X_i = 1]$.
 - (c) Muestre que $P\{S > x\} = \lambda E[I(S > x)/S | X_i = 1]$, donde $I(S > x)$ es igual a 1 si $S > x$ e igual a cero en caso contrario.
11. Suponga que se dispone de un conjunto de n componentes, en el que el componente j funciona (de manera independiente) con probabilidad p_j , $j = 1, \dots, n$. Queremos realizar m experimentos. Sin embargo, para realizar el experimento i ,

deben funcionar todos los componentes del conjunto \mathcal{A}_i , $i = 1, \dots, m$. Si se puede realizar el experimento i , entonces ganamos una cantidad a_i . Si la ganancia total es la suma de las ganancias de los m experimentos, suponga que estamos interesados en estimar la probabilidad de que la ganancia total sea mayor que x . Suponiendo que esta probabilidad es pequeña, dé un procedimiento eficaz de simulación para estimarla.

12. Sea Z_1, Z_2, \dots una sucesión de normales estándar independientes. Sea

$$M_n = \frac{Z_n + Z_{n-1} + Z_{n-2} + Z_{n-3}}{4}, \quad n \geq 4$$

y definamos N como

$$N = \text{Mín} \left\{ n: |M_n| > \frac{3}{2} \right\}$$

Queremos emplear simulación para determinar $E[N]$. Determine la varianza del estimador de simulación bruto y luego la de aquél que utiliza la suma de los riesgos aleatorios como control (lo anterior es un diagrama de control con promedio móvil que sirve para decidir si la distribución se ha desfasado de la normal estándar).

13. Sean X_1, X_2, \dots variables aleatorias geométricas independientes con parámetro $p = \frac{1}{100}$. Sea

$$N = \text{mín} \{ n: X_n = X_k, \text{ para algunas } k = 1, \dots, n-1 \}$$

el primer instante en que una variable aleatoria tiene un valor igual al de uno de sus predecesores.

- (a) Utilice simulación para estimar $\text{Var}(N)$.
- (b) Estime la varianza del estimador de $E[N]$ que utiliza la suma de los riesgos aleatorios como variable de control.

Bibliografía

- El Khadiri, M. y H. Cancela, "An Improvement to the Total Hazard Method for System Reliability Simulation", en *Probability in the Engineering and Informational Sciences*, **10**, 1996, p. 2.
- Kronmal, R. A. y A. V. Peterson, Jr., "On the Alias Method for Generating Random Variables from a Discrete Distribution", en *American Statistician*, **33**, 1979, pp. 214-218.

- Peterson, A. V., Jr. y R. A. Kronmal, "On Mixture Methods for the Computer Generation of Random Variables", en *American Statistician*, **36**, 1982, pp. 184-191.
- Ross, S. M., "Variance Reduction in Simulation via Random Hazards", en *Probability in the Engineering and Informational Sciences*, **4**, 1990, pp. 299-310.
- Ross, S. M., "A New Simulation Estimator of System Reliability", en *Journal of Applied Mathematics and Stochastic Analysis*, **7**, 1994, pp. 331-336.

Apéndice de programas

- 4-1 Simulación de variables aleatorias Poisson
- 9-0 La función de distribución normal estándar
- 9-1 La función de distribución ji-cuadrada
- 9-2 Uso de la simulación para aproximar el valor p en la prueba de bondad de ajuste
- 9-3 Uso de la simulación para aproximar el valor p en la prueba de Kolmogorov-Smirnov
- 9-4 El valor p exacto para la prueba de la suma de rangos de dos muestras
- 9-5 Uso de la simulación para aproximar el valor p en la prueba de suma de rangos de dos muestras

Programa 4-1

```
10 PRINT "ESTE PROGRAMA SIMULARÁ VARIABLES ALEATORIAS
    POISSON INDEPENDIENTES E IDÉNTICAMENTE DISTRIBUIDAS"
20 PRINT "PROPORCIONE LA MEDIA"
30 INPUT L
40 PRINT "PROPORCIONE N"
50 INPUT N
60 RANDOMIZE
70 I = INT (L)
80 B = 1/L
90 FOR K = 1 TO I
100 S = S + LOG (K)
110 NEXT
120 S = -S -L + I * LOG (L)
```

Apéndice de programas

```
130 PP = EXP (S)
140 F = 1
150 FOR K = 1 TO I
160 F = F* (I + 1-K) * B
170 CUM = CUM + F
180 NEXT
190 CUM = (CUM + 1) * PP
200 FOR J = 1 TO N
210 I = INT (L)
220 A = CUM
230 P = PP
240 U = RND
250 IF U < A GOTO 310
260 I = I + 1
270 P = L * P/I
280 A = A + P
290 IF U < A THEN GOTO 360
300 GOTO 260
310 A = A-P
320 IF U > A GOTO 360
330 P = I * P * B
340 I = I - 1
350 GOTO 310
360 PRINT I
370 NEXT
380 END
```

Programa 9-0

```
10 PRINT "ESTE PROGRAMA CALCULA LA PROBABILIDAD DE QUE UNA
    VARIABLE ALEATORIA NORMAL ESTÁNDAR SEA MENOR QUE X"
20 PRINT "PROPORCIONE EL VALOR DESEADO DE X"
30 INPUT X
40 U = ABS (X)
50 IF U > 4 GOTO 180
60 Y = U ^ 2
70 I = U
80 FOR J = 1 TO 40
90 U = -U * Y * (2 * J - 1)/(2 * J * (2 * J + 1))
```

```

100 I = I + U
110 NEXT
120 I = I/SQR (2 * 3.14159)
130 IF X < 0 GOTO 160
140 PRINT "LA PROBABILIDAD ES"; .5 + I
150 GOTO 220
160 PRINT "LA PROBABILIDAD ES"; .5 - I
170 GOTO 220
180 IF X < 0 GOTO 210
190 PRINT "LA PROBABILIDAD ES MAYOR QUE"; 1 - 10^-4
200 GOTO 220
210 PRINT "LA PROBABILIDAD ES MENOR QUE"; 10^-4
220 END

```

Programa 9-1

```

10 PRINT "ESTE PROGRAMA CALCULA LA PROBABILIDAD DE QUE UNA
    VARIABLE ALEATORIA JI CUADRADA CON N GRADOS DE LIBERTAD
    SEA MENOR QUE X"
20 PRINT "PROPORCIONE EL PARÁMETRO DE GRADOS DE LIBERTAD"
30 INPUT N
40 S = (N - 1)/2
50 PRINT "PROPORCIONE EL VALOR DESEADO DE X"
60 INPUT X
70 M = X/2
80 D = X/2 - N/2 + 1 /3
90 D = D - .04/N
100 IFN = 1 GOTO 160
110 IFS = M GOTO 180
120 H = S/M
130 X = (1 - H * H + 2 * H * LOG (H)/(1 - H)^ 2
140 X = D * SQR ((1 + X)/M)
150 GOTO 190
160 X = D * SQR (2/M)
170 GOTO 190
180 X = D/SQR (M)
190 U = ABS (X)
200 IF U > 4 GOTO 330
210 Y = U ^ 2

```

```

220 I = U
230 FOR J = 1 TO 40
240 U = -U * Y * (2 * J - 1)/(2 * J * (2 * J + 1))
250 I = I + U
260 NEXT
270 I = I/SQR (2 * 3.14159)
280 IF X < 0 GOTO 310
290 PRINT "LA PROBABILIDAD ES"; .5 + I
300 GOTO 370
310 PRINT "LA PROBABILIDAD ES"; .5 - I
320 GOTO 370
330 IF X < 0 GOTO 360
340 PRINT "LA PROBABILIDAD ES MAYOR QUE"; 1 - 10^-4
350 GOTO 370
360 PRINT "LA PROBABILIDAD ES MENOR QUE"; 10^-4
370 END

```

Programa 9-2

```

10 PRINT "ESTE PROGRAMA UTILIZA SIMULACIÓN PARA APROXIMAR
    EL VALOR p EN LA PRUEBA DE BONDAD DE AJUSTE"
20 RANDOMIZE
30 PRINT "PROPORCIONE EL NÚMERO DE POSIBLES VALORES"
40 INPUT N
50 DIM P (N)
60 DIM Q (N)
66 Q (0) = 0
70 PRINT "PROPORCIONE LAS PROBABILIDADES UNA A LA VEZ"
80 FOR I = 1 TO N
90 INPUT P (I)
100 Q (I) = Q (I - 1) + P (I)
110 NEXT I
120 PRINT "PROPORCIONE EL TAMAÑO DE LA MUESTRA"
130 INPUT D
140 DIM B (N)
150 DIM C (N)
160 FOR J = 1 TO N
170 B (J) = D * P (J)
180 C (J) = 1/B (J)

```

```

190 NEXT
200 DIM X(N)
210 PRINT "PROPORCIONE LA CANTIDAD DESEADA DE EJECUCIONES
      DE SIMULACIONES"
220 INPUT R
230 PRINT "PROPORCIONE EL VALOR DEL ESTADÍSTICO DE PRUEBA"
240 INPUT W
250 FOR K = 1 TO R
260 FOR L = 1 TO N
270 X (L) = 0
280 NEXT L
290 FOR J = 1 TO D
300 U = RND
310 I = 1
320 IF U < Q (I) THEN GOTO 350
330 I = I + 1
340 GOTO 320
350 X (I) = X (I) + 1
360 NEXT J
370 S = 0
380 FOR L = 1 TO N
390 S = S + (X(L) - B(L)) ^ 2 * C (L)
400 NEXT L
410 IF S > = W THEN C = C + 1
420 NEXT K
430 PRINT "LA ESTIMACIÓN DEL VALOR p ES"; C/R
440 END

```

Programa 9-3

```

10 PRINT "ESTE PROGRAMA UTILIZA SIMULACIÓN PARA APROXIMAR
      EL VALOR p DE LA PRUEBA DE KOLMOGOROV-SMIRNOV"
20 RANDOMIZE
30 PRINT "PROPORCIONE EL VALOR DE LA CANTIDAD DE PRUEBA"
40 INPUT D
50 PRINT "PROPORCIONE EL TAMAÑO DE LA MUESTRA"
60 INPUT N
70 A = 1/N
80 PRINT "PROPORCIONE EL NÚMERO DESEADO DE EJECUCIONES DE
      SIMULACIÓN"

```

```

90 INPUT R
100 DIM T (N)
110 FOR I = 1 TO R
120 S = 0
130 T (0) = 0
140 FOR J = 1 TO N
150 U = RND
160 Y = -LOG (U)
170 T (J) = T (J - 1) + Y
180 S = S + Y
190 NEXT
200 S = S - LOG (RND)
210 C = 1/S
220 J = 1
230 X = T (J) * C
240 IF J * A - X > D THEN GOTO 290
250 IF X - (J - 1) * A > D THEN GOTO 290
260 IF J = N GOTO 300
270 J = J + 1
280 GOTO 230
290 CC = CC + 1
300 NEXT I
310 PRINT "EL VALOR p APROXIMADO ES"; CC/R
320 END

```

Programa 9-4

```

10 PRINT "ESTE PROGRAMA CALCULA EL VALOR p PARA LA PRUEBA DE
      SUMA DE RANGOS DE DOS MUESTRAS"
20 PRINT "ESTE PROGRAMA SERÁ MÁS RÁPIDO SI USTED DESIGNA
      COMO PRIMERA MUESTRA AQUELLA CON MENOR SUMA DE RANGOS"
30 PRINT "PROPORCIONE EL TAMAÑO DE LA PRIMERA MUESTRA"
40 INPUT N
50 PRINT "PROPORCIONE EL TAMAÑO DE LA SEGUNDA MUESTRA"
60 INPUT M
70 PRINT "PROPORCIONE LA SUMA DE LOS RANGOS DE LA PRIMERA
      MUESTRA"
80 INPUT T
90 DIM P(N,M,T + 1)

```

```

100 FOR I = 1 TO N
110 FOR K = I * (I + 1)/2 TO T
120 P (I, 0, K) =1
130 NEXT
140 NEXT
150 FOR K = 1 TO T + 1
160 FOR J = 1 TO M
170 P (0, J, K-1) =1
180 NEXT
190 NEXT
200 FOR I = TO N
210 FOR J = 1 TO M
220 FOR K = 1 TO T
230 IF K < (I + J) THEN P(I, J, K) = (J/(I + J)) * P(I,
    J - 1, K) ELSE P(I, J, K) = (I/(I + J)) * P(I - 1, J,
    K - I - J) + (J/(I + J)) * P(I, J - 1, K)
240 NEXT
250 NEXT
260 NEXT
270 IF P(N, M, T) < 1-P(N, M, T-1) THEN V = P(N, M, T) ELSE
    V = 1 - P(N, M, T - 1)
280 PRINT "EL VALOR p ES"; 2*V
290 END

```

Programa 9-5

```

10 PRINT "ESTE PROGRAMA APROXIMA EL VALOR p EN LA PRUEBA DE
    SUMA DE RANGOS DE DOS MUESTRAS MEDIANTE UN ESTUDIO DE
    SIMULACIÓN"
20 RANDOMIZE
30 PRINT "PROPORCIONE EL TAMAÑO DE LA PRIMERA MUESTRA"
40 INPUT N(1)
50 PRINT "PROPORCIONE EL TAMAÑO DE LA SEGUNDA MUESTRA"
60 INPUT N(2)
70 PRINT "PROPORCIONE LA SUMA DE LOS RANGOS DE LA PRIMERA
    MUESTRA"
80 INPUT T
90 PRINT "PROPORCIONE EL NÚMERO DESEADO DE EJECUCIONES DE
    LA SIMULACIÓN"

```

```

100 INPUT M
110 N = N(1) + N(2)
120 DIM X (N)
130 NUM = NUM + 1
140 S = 0
150 FOR I = 1 TO N
160 X (I) = I
170 NEXT I
180 FOR I = 1 TO N (1)
190 R = INT ((N + 1 - I) * RND) + 1
200 S = S + X(R)
210 X (R) = X(N + 1 - I)
220 NEXT I
230 IF S <= T THEN C(1) = C(1) + 1
240 IF S >= T THEN C(2) = C(2) + 1
250 IF NUM < M GOTO 130
260 IF C(1) > C(2) THEN C(1) = C(2)
270 PRINT "EL VALOR p APROXIMADO ES"; 2*C(1)/M
280 END

```

Índice

A

Aleatoria, variable, 8
 continua, 8-9
 discreta, 8
Algoritmo
 de adelgazamiento, 78
 de Hastings-Metropolis, 221-224
 de la transformada inversa,
 para generar variables aleatorias
 continuas, 62
 de muestreo con remuestreo de
 importancia, 239-242
 para generar variables aleatorias
 discretas, 45-46
Axiomas de la probabilidad, 6

B

Blackjack, 146-147

C

Cadena de Markov, 218-221
 aperiódica, 220

irreducible, 219
 probabilidades estacionarias de, 219
 reversible con respecto del tiempo,
 220

Contadores, 87
Control de calidad, 132
Covarianza, 14
Criterio de suma de rangos de varios
 muestreos, 209

D

Desigualdad
 de Chebyshev, 16
 de Markov, 15
Desviación estándar muestral, 115
Diagrama de control de suma
 acumulativa, 172, 262-263

E

Energía Ising de una matriz, 244
Error cuadrático medio, 112
Esperanza condicional, 30
 en la reducción de la varianza,
 149-150

Estadística bayesiana, 241-242

Estimación

condicional de intervalos, 119

de e , 129, 139-141

de π , 40-42, 150-152, 163

Estimador insesgado, 112

Estructura

de puente, 135

en serie, 134

k de n , 135

paralela, 135

Eventos independientes, 8

F

Fórmula para la varianza condicional, 31

Función

de densidad de probabilidad, 9

conjunta, 10

de distribución, 8

conjunta, 9

empírica, 12

de fiabilidad, 136, 143, 152-153, 257-258

de masa de probabilidad, 8

conjunta, 10

L

Ley

débil de los grandes números, 16

fuerte de los grandes números, 17

Línea de espera (cola), 137, 144,

154-155, 171-172

capacidad finita, 153-154

periodo ocupado, 263-264

servidor paralelo, 94-97

servidor único, 4, 88-90

tándem (o secuencial), 90, 92-94

Lista de eventos, 87

M

Media muestral, 112

cálculo recursivo, 116

Media. Véase Valor esperado.

Medias por lotes, 221

Método

aplicado a un sistema de línea de espera, 125-127

bootstrap en estadística, 121-124

congruencial mixto, 37

congruencial multiplicativo, 37

de composición, 56-57

de la transformada inversa discreta, 45-46

de Monte Carlo, 38

de Monte Carlo para una cadena de Markov, 220-221

de rechazo, 53-56, 66-67

de regeneración, 127-128

de simulación por medio de eventos discretos, 86-87

de tasa de riesgo discreta, 60

de variables alias, 248-252

dentro-fuera (hit-miss), 184-185

polar, 72-75

Modelo

de inventario, 97-99

de opción de acciones, 101, 102-104

de probabilidad, 2

de reparación de una máquina, 99-101

jerárquico de Bayes, 234

Muestreador de Gibbs, 224-236

Muestreo

de importancia, 167-168

estratificado en la reducción de la varianza, 158-162, 266

Mutualmente excluyentes, 6

N

Número(s)

aleatorio, 36

aleatorios comunes, 178-179

pseudoaleatorios, 36

P

Patrones, 259-260

Permutación aleatoria, 47-48, 58, 173

Probabilidad condicional, 7

Problema

de dos muestras, 202-203

de reordenamiento de una lista, 145-146

del agente de ventas viajero, 238

Proceso

de renovación, 156

bidimensional, 252-255

no estacionario. Véase Proceso

no homogéneo, 29, 77-81, 87

Poisson, 26-29, 75-77

Poisson no homogéneo

Prueba de la suma del rango de dos muestras, 202-204

valor p , 204-208

Pruebas de bondad de ajuste, 189-190

de Kolmogorov-Smirnov, 193-198, 201-202

χ^2 cuadrada, 190-191, 198-199

R

Redes de colas, 226-228, 229-231

Regla de control del promedio con movimiento de peso exponencial, 131-133

Riesgos, 260

aleatorios, 260-262

T

Tasas de riesgo discretas, 60

Técnica de aceptación y rechazo. Véase

Método de rechazo

Temple simulado, 236-238

Teorema central del límite, 25

Transformaciones de Box-Muller, 73

Transformada de Escher de una densidad (*tilted density*), 168, 169

V

Validación, 189

de una hipótesis de proceso Poisson, 212-214

de una hipótesis de proceso Poisson no homogéneo, 210-212

Valor esperado, 10-13

Valor p , 191

Variable(s) aleatoria(s)

Bernoulli, 17-18, 117, 120, 256

beta, 67

binomial, 17-18, 33, 34, 52-53

binomial negativa, 21, 59

continua, 8-9

de Cox, 84

exponencial, 25-26, 28, 35, 63-64, 65-66

gamma, 28-29, 65, 68

geométrica, 20, 49-50

hipergeométrica, 21

independientes, 10

normales, 23-25, 70-71

estándar, 24, 59

unitaria. Véase normal estándar

normal unitaria. Véase Variable

aleatoria normal estándar

Poisson, 19, 34, 50-52, 64



AGG

IMP. Y MAQ. DE LIBROS MIG, S.A. DE C.V.
VENADO No. 530
COL. LOS OLIVOS, TLAHUAC
MEXICO 13210, D.F.

1999

