

Estadística Computacional

Estadística Computacional

Antonio Salmerón Cerdán

María Morales Giraldo

ALMERÍA, 2001

Prólogo

El objetivo de este trabajo es ofrecer un libro de texto que cubra los contenidos teóricos de las asignaturas de Estadística Computacional que se imparten en la Licenciatura en Matemáticas y en las Ingenierías Técnicas en Informática de Gestión y de Sistemas.

Se ha buscado un compromiso entre rigurosidad y claridad de exposición, de manera que el libro sea accesible tanto a los alumnos de Informática como a los de Matemáticas.

La mayoría de las técnicas contenidas en este texto han sido presentadas con su motivación y desarrollo teórico acompañado de ejemplos aclaratorios y ejercicios de comprensión. Se han especificado también numerosos algoritmos en pseudo-código, fácilmente traducibles a un lenguaje de propósito general (como C o Pascal) u orientados a la Estadística (como R).

Los alumnos de Informática pueden prescindir de las demostraciones y centrarse en los aspectos algorítmicos.

Índice General

1	Introducción	13
1.1	Consideraciones generales sobre la Estadística Computacional	13
1.2	Organización por capítulos	15
2	Generación de números aleatorios	17
2.1	Definición y primeros métodos de generación	17
2.1.1	Métodos manuales	18
2.1.2	Tablas y mecanismos analógicos	18
2.2	Métodos digitales	19
2.2.1	Generadores congruenciales	20
2.2.1.1	Generador congruencial lineal	20
2.2.1.2	Generador Congruencial Multiplicativo (G.C.M.)	30
2.2.1.3	Otros generadores congruenciales	31
2.3	Métodos de mejora de la aleatoriedad	33
2.3.1	Método aditivo de dos series	33
2.3.2	Algoritmo de mezcla de dos series	33
2.3.3	Algoritmo de mezcla con una sola serie	34

3	Comprobación de la aleatoriedad	37
3.1	Test χ^2	37
3.2	Test de series	40
3.3	Test de rachas	42
3.4	Test de huecos o de distancias	44
3.5	Test de Kolmogorov-Smirnov	45
4	Generación de muestras	49
4.1	Método de inversión	50
4.1.1	Método de inversión para variables discretas	53
4.2	Método de composición	54
4.3	Método de aceptación-rechazo	60
4.4	Métodos particulares	67
4.4.1	Distribuciones Exponencial, Gamma y χ^2	67
4.4.2	Distribución Normal	69
4.4.2.1	Uso del teorema central del límite	70
4.4.2.2	Método de Box-Muller	71
4.4.2.3	Método de aceptación-rechazo para la distribución Normal	72
4.4.3	Distribución Binomial	74
4.4.4	Distribución de Poisson	74
5	Integración Monte Carlo	77
5.1	Método de ensayo-error	77
5.2	Método de la media muestral	84

5.3	Eficiencia del método de Monte Carlo	92
5.4	Mejora de la eficiencia	93
5.4.1	Muestreo por importancia	93
5.4.2	Muestreo estratificado	102
5.4.3	Muestreo sistemático	110
5.4.4	Variables antitéticas	112
6	Técnicas de remuestreo	115
6.1	Remuestreo	115
6.1.1	Error estándar de la media	116
6.2	Bootstrap	117
6.3	Jackknife	120
	Bibliografía	123

Capítulo 1

Introducción

1.1 Consideraciones generales sobre la Estadística Computacional

Puede definirse la *Estadística Computacional* como la Estadística calculada. Desde este punto de vista, cualquier procedimiento estadístico que implique una gran cantidad de cálculos, puede considerarse como un procedimiento de Estadística Computacional.

Una de las ramas más destacadas dentro de la Estadística Computacional es la correspondiente a la *simulación y métodos de Monte Carlo*. Las técnicas estadísticas de simulación juegan un papel muy importante en las ciencias experimentales e ingenierías, como la Informática, Física, Química, Biología y, naturalmente, dentro de la propia Estadística.

La *simulación* es una técnica numérica para realizar experimentos en un ordenador, sobre cierto tipo de modelos matemáticos, de cara a hacer predicciones sobre el comportamiento de dichos modelos.

Podemos definir el *método de Monte Carlo* como una técnica para realizar experimentos (con la ayuda de un ordenador) sobre un determinado modelo de cara a obtener información o realizar afirmaciones sobre los valores de ciertos parámetros de dicho modelo. En este contexto, entenderemos por *modelo* un conjunto de variables aleatorias y una serie de relaciones entre ellas.

La presencia de variables aleatorias conlleva un proceso de muestreo durante la experimentación, en el que se utilizan números aleatorios para generar valores de las variables en cuestión.

La denominación de Monte Carlo se debe al nombre clave que usó von Neumann para referirse al proyecto de investigación que desarrollaba en el centro de Los Alamos en la Segunda Guerra Mundial, relativa al desarrollo de una bomba atómica. Simulaban el comportamiento de la ecuación de Boltzman, que modeliza la difusión aleatoria de los neutrones en el material fisionable. Al utilizar números aleatorios en la simulación, recurrieron al término Monte Carlo en alusión a los casinos de la ciudad del mismo nombre en Mónaco.

Pero el método de Monte Carlo también puede emplearse para abordar problemas de carácter determinista, para lo cual será necesario representar el problema a tratar mediante un modelo aleatorio. En este sentido, el objetivo de introducir un componente aleatorio es ayudar a simplificar el modelo, aunque a cambio de ello, en lugar de obtener resultados exactos obtendremos aproximaciones o estimaciones de los resultados exactos.

Figura 1.1: Queremos estimar el área del círculo

Un ejemplo de problema determinista que se presta a la aplicación de una técnica de Monte Carlo es el cálculo de una integral definida.

Por ejemplo, podríamos estimar el área del círculo de la figura 1.1 generando puntos al azar dentro del cuadrado y calculando luego la proporción de puntos que han acertado dentro del círculo.

La forma de operar de un método de Monte Carlo está representada en la figura 1.2. En primer lugar, es necesario disponer de un mecanismo generador de números



Figura 1.2: Esquema de la simulación Monte Carlo

aleatorios. A partir de los números que genera este mecanismo, se obtiene una muestra que es utilizada para estimar los valores de los parámetros de interés.

1.2 Organización por capítulos

Comenzaremos estudiando la generación de números aleatorios en el capítulo 2. La validez de una secuencia candidata de números aleatorios deberá ser contrastada mediante las llamadas pruebas de aleatoriedad, a las que dedicamos el capítulo 3. En el capítulo 4 veremos cómo obtener una muestra artificial a partir de una distribución de probabilidad genérica supuesto que disponemos de un mecanismo generador de números aleatorios. Particularizaremos al caso de algunas distribuciones notables. En el capítulo 5 estudiaremos el método de Monte Carlo resolviendo un problema determinista muy general: el cálculo de integrales definidas. Terminamos el libro estudiando brevemente, en el capítulo 6 las técnicas de remuestreo, que permiten mejorar la calidad de una muestra.

Capítulo 2

Generación de números aleatorios

Los números aleatorios son de gran utilidad en distintas facetas de las ciencias experimentales y también en áreas como la Informática. Desde un punto de vista general, los números aleatorios constituyen las entradas de un modelo de simulación, y como tales, el funcionamiento de la simulación depende de ellos. En este capítulo discutiremos en cierta profundidad algunos mecanismos para la obtención de secuencias de números aleatorios.

2.1 Definición y primeros métodos de generación

En realidad, hablar de que un número de por sí sea o no aleatorio no parece acertado. Más bien, tiene sentido hablar de si una secuencia de números es o no aleatoria.

Definición 2.1 *Una secuencia de números aleatorios es una sucesión de variables aleatorias $\{X_1, X_2, \dots, X_n\}$ donde:*

1. *Cada variable X_i sigue una distribución uniforme en el intervalo $[0, 1)$.*
2. *Las variables $\{X_1, X_2, \dots, X_n\}$ son independientes.*

Por lo tanto, la aleatoriedad la consideraremos caracterizada por las dos propiedades de la definición anterior: frecuencias e independencia, respectivamente. Nuestro objetivo es buscar métodos para obtener de forma automática secuencias de números

aleatorios, concretamente, secuencias que podamos usar en programas de simulación. Desde este punto de vista, un buen método deberá caracterizarse por:

1. *Reproductividad*: Toda secuencia de números aleatorios debe poderse reproducir con objeto de que los experimentos realizados a partir de ella sean contrastables por otro experimentador.
2. *Velocidad*: El coste temporal es una variable a considerar, dado que los números aleatorios son tan sólo el inicio del proceso de simulación, que ya de por sí suele ser muy costoso.
3. *Economía* en cuanto a requerimientos de memoria, por la misma razón que en el caso anterior.

Podemos a clasificar los métodos de generación de números aleatorios en cuatro grupos: *Métodos manuales, tablas, analógicos y digitales*.

2.1.1 Métodos manuales

Cuando los investigadores comenzaron a necesitar números aleatorios en sus trabajos, recurrieron a métodos como el lanzamiento de dados, monedas, extracción de cartas y procedimientos similares cuyas limitaciones son obvias, y su utilidad prácticamente nula dentro del contexto de la Estadística Computacional.

2.1.2 Tablas y mecanismos analógicos

En 1927, L.H.C. Tippett publicó la primera tabla con 40.000 números aleatorios obtenida aleatoriamente a partir del censo. Desde entonces fueron sucediéndose distintos dispositivos de generación de números aleatorios; el primero de ellos, diseñado por M.G. Kendall y B. Babington-Smith en 1939, construyó una secuencia de 100.000 números aleatorios. En 1955 se consiguieron secuencias de un millón de números aleatorios y posteriormente se utilizó un dispositivo denominado ERNIE para extraer los números ganadores de la lotería inglesa.

Pero con la introducción de los ordenadores el uso de las tablas hacía los procesos muy lentos y ocupaban mucho espacio en memoria. Además, los dispositivos como ERNIE producían secuencias imposibles de generar por segunda vez.

Esto provocó la necesidad de diseñar algoritmos numéricos que pudiesen ser implementados en un ordenador, a estos algoritmos les llamamos métodos digitales. Dedicaremos el resto del capítulo a estos métodos, ya que son los que se utilizan actualmente al proporcionar de forma sencilla los números necesitados por cualquier programa de simulación.

2.2 Métodos digitales

Los métodos digitales son algoritmos numéricos que producen secuencias de números X_1, \dots, X_n con $X_{n+1} = f(X_n)$.

Obviamente, al depender cada número de su predecesor mediante una dependencia funcional, la secuencia deja de ser aleatoria; sin embargo, es posible obtener secuencias de tal calidad que las pruebas estadísticas de aleatoriedad no sospechen de las mismas.

Definición 2.2 *A los números de una secuencia obtenida determinísticamente los llamaremos números pseudoaleatorios.*

A partir de este momento, cuando no haya posibilidad de confusión, seguiremos hablando de números aleatorios aunque realmente nos referimos a números pseudoaleatorios.

El primer método de generación de números aleatorios en ordenador fue el llamado *Método del Centro de los Cuadrados*, propuesto por John Von Neumann en 1946. Este método consiste en tomar un número aleatoriamente y generar la secuencia elevando al cuadrado las cifras centrales del número predecesor.

Ejemplo 2.1 *En la siguiente tabla se muestra una secuencia de números aleatorios generados mediante el método del centro de los cuadrados de Von Neumann:*

n	n^2	X_i
23	0 52 9	0.52
52	2 70 4	0.70
70	4 90 0	0.90
90	8 10 0	0.10
10	0 10 0	0.10
10	0 10 0	0.10
...

En el ejemplo podemos observar el principal problema de este método: podemos llegar a un número en el que, a partir de él, todos los números sean iguales; por ejemplo, si el cero aparece como número de la secuencia, el resto de la secuencia estará formada por ceros.

2.2.1 Generadores congruenciales

2.2.1.1 Generador congruencial lineal

El generador congruencial lineal (G.C.L.) fue propuesto por D.H. Lehmer en 1949 y consiste en, a partir de un número inicial llamado *semilla*, generar la secuencia por recurrencia según el esquema:

$$X_{n+1} = (aX_n + c) \pmod{m}, \quad (2.1)$$

en donde

$$\begin{aligned} a, & \text{ es el } \textit{multiplicador} & (0 \leq a < m) \\ m, & \text{ el } \textit{módulo} & (m > 0) \\ c, & \text{ el } \textit{incremento} & (0 \leq c < m) \\ X_0, & \text{ la } \textit{semilla} & (0 \leq X_0 < m) \end{aligned}$$

Aunque técnicamente es posible tomar los valores $a = 0$ o $a = 1$, en la práctica no parece que obtengamos buenos resultados, pues

- si $a \neq 0$ obtendríamos siempre el mismo número,

- si $a = 1$ entonces, al aplicar la fórmula (2.1) tenemos $X_{n+1} = X_n + c$ esto es:

$$\begin{aligned} X_1 &= X_0 + c \\ X_2 &= X_1 + c = X_0 + 2c \\ &\vdots \\ X_n &= X_0 + nc \end{aligned}$$

siendo razonable sospechar de la aleatoriedad de esta secuencia.

Ejemplo 2.2 *Veamos un ejemplo de una secuencia obtenida mediante un G.C.L.: si tomamos $X_0 = a = c = 3$ y $m = 5$, obtenemos la secuencia*

$$3, 2, 4, 0, 3, 2, 4, 0, 3, 2, 4, 0, \dots$$

Este ejemplo pone en relieve los siguientes aspectos:

1. Las secuencias obtenidas se acercarán más o menos a la aleatoriedad dependiendo de los valores de m , a , c y X_0 . Posteriormente veremos cómo elegir adecuadamente estos valores.
2. La ecuación (2.1) produce números entre 0 y $m - 1$, por lo que debemos dividir por m para llevarlos al intervalo $[0, 1)$. La secuencia del ejemplo anterior, una vez dividida por m , sería:

$$0.6, 0.4, 0.8, 0, 0.6, 0.4, 0.8, 0, 0.6, 0.4, 0.8, 0, \dots$$

3. Las secuencias de números obtenidas a través de un generador de la forma $X_{n+1} = f(X_n)$ siempre son cíclicas, repitiéndose el ciclo, llamado **periodo**, indefinidamente. El periodo máximo de un G.C.L. será como mucho de m términos, a partir del m -ésimo elemento se repite exactamente la misma secuencia.
4. A la longitud de la secuencia máxima que se pueda obtener con un G.C.L. se le llama **longitud de periodo**. Obviamente, intentaremos conseguir secuencias con la mayor longitud de periodo posible.

Dada la forma recursiva en que se generan los números en un G.C.L., éstos tienen algunas propiedades que hacen que la secuencia sea predecible. Por ejemplo, podemos obtener la secuencia que se obtiene a partir de otra cogiendo solamente un número de entre k . Es decir, partimos de $x_{n+1} = (ax_n + c) \pmod m$ y queremos obtener la sucesión $y_0 = x_0, y_1 = x_k, y_2 = x_{2k}, \dots$, con $k =$ longitud del salto. En general, queremos obtener la sucesión $y_i = x_{ki}$. La siguiente proposición nos dice cómo.

Proposición 2.1 *Sea x_n la secuencia obtenida a partir de un G.C.L., y sea $y_i = x_{ki}$, entonces,*

$$y_{n+1} = \left(a^k y_n + \frac{a^k - 1}{a - 1} \cdot c \right) \pmod m$$

Demostración: Sabemos que $y_{n+1} = x_{(n+1)k} = x_{kn+k}$. Por otro lado, $y_n = x_{kn}$. Avanzando desde aquí, veremos qué expresión tiene x_{kn+k} .

$$\begin{aligned} x_{kn+1} &= (ax_{kn} + c) \pmod m \\ x_{kn+2} &= (ax_{kn+1} + c) \pmod m = [a((ax_{kn} + c) \pmod m) + c] \pmod m \\ x_{kn+3} &= (ax_{kn+2} + c) \pmod m \\ &= [a(a(((ax_{kn} + c) \pmod m) + c) \pmod m) + c] \pmod m \\ &\vdots \\ x_{kn+k} &= (a^k x_{kn} + a^{k-1}c + a^{k-2}c + \dots + c) \pmod m \\ &= [a^k x_{kn} + (a^{k-1} + a^{k-2} + \dots + 1)c] \pmod m \\ &= \left(a^k x_{kn} + \frac{a^k - 1}{a - 1} \cdot c \right) \pmod m \\ &= \left(a^k y_n + \frac{a^k - 1}{a - 1} \cdot c \right) \pmod m \end{aligned}$$

□

Una consecuencia de este resultado es que, si estamos generando una secuencia, en cada momento podemos saber qué número saldrá dentro de k posiciones.

Veamos ahora cómo elegir valores apropiados para los parámetros del modelo:

Elección del módulo

Para la elección del módulo tendremos en cuenta dos factores:

- m debe ser lo mayor posible pues el periodo será menor o igual que el módulo.
- El otro factor determinante es la velocidad de generación.

El cálculo de $x \pmod{y}$ requiere el uso de la división, operación lenta de ejecutar en el ordenador, por ello es conveniente escoger $m = 2^r$ siendo r el tamaño de palabra del ordenador. Con esto se evita el uso de la división pues el cálculo del módulo se reduce a considerar las últimas r cifras del número.

Elección del multiplicador

De la secuencia generada, deberíamos esperar que contenga el mayor número de dígitos posibles, por lo tanto, ahora nuestro objetivo será escoger el valor de a de forma que maximice la longitud del periodo de un G.C.L.

Comenzaremos demostrando en primer lugar unos resultados previos que usaremos para probar el teorema que caracteriza la elección de los parámetros.

Lema 2.1 *Sea p un número primo, y sea e un entero positivo con $p^e > 2$. Si*

$$x \equiv 1 \pmod{p^e}, x \not\equiv 1 \pmod{p^{e+1}} \quad (2.2)$$

entonces

$$x^p \equiv 1 \pmod{p^{e+1}}, x \not\equiv 1 \pmod{p^{e+2}}. \quad (2.3)$$

Demostración: $x \equiv 1 \pmod{p^e}$ significa por definición que x y 1 tienen el mismo resto al ser divididos por p^e , es decir,

$$x = qp^e + 1 \text{ con } q \text{ entero}$$

y para que $x \not\equiv 1 \pmod{p^{e+1}}$, q no debe ser múltiplo de p .

$$\begin{aligned}
x^p &= (qp^e + 1)^p \\
&= 1 + \binom{p}{1} qp^e + \binom{p}{2} (qp^e)^2 + \cdots + \binom{p}{p-1} (qp^e)^{p-1} + \binom{p}{p} (qp^e)^p \\
&= 1 + qp^{e+1} + \binom{p}{2} q^2 p^{2e} + \cdots + \binom{p}{p-1} q^{p-1} p^{(p-1)e} + q^p p^{pe} \\
&= 1 + qp^{e+1} \underbrace{\left(1 + \binom{p}{2} qp^{e-1} + \cdots + \binom{p}{p-1} q^{p-2} p^{(p-2)e-1} + q^{p-1} p^{(p-1)e-1} \right)}_{\in \mathbb{Z}}
\end{aligned}$$

Luego $x^p \equiv 1 \pmod{p^{e+1}}$ y como el término entre paréntesis no es múltiplo de p , $x \not\equiv 1 \pmod{p^{e+2}}$ \square

Lema 2.2 Sea $m = p_1^{e_1} \cdot p_2^{e_2} \cdot \cdots \cdot p_n^{e_n}$ la descomposición de m en números primos. La longitud λ del periodo de la secuencia determinada por un G.C.L. de parámetros (X_0, a, c, m) es el mínimo común múltiplo de las longitudes λ_i , siendo λ_i la longitud del periodo de la secuencia generada por el G.C.L. de parámetros $(X_0 \pmod{p_i^{e_i}}, a \pmod{p_i^{e_i}}, c \pmod{p_i^{e_i}}, p_i^{e_i})$, $1 \leq i \leq n$.

Demostración: Probemos primero el caso en que $m = m_1 \cdot m_2$:

- Sean $\{X_n\}$ los elementos de la secuencia generada con los parámetros (X_0, a, c, m) ,
- sean $\{Y_n\}$ los elementos de la secuencia generada con los parámetros $(X_0 \pmod{m_1}, a \pmod{m_1}, c \pmod{m_1}, m_1)$,
- sean $\{Z_n\}$ los elementos de la secuencia generada con los parámetros $(X_0 \pmod{m_2}, a \pmod{m_2}, c \pmod{m_2}, m_2)$

Entonces, $Y_n = X_n \pmod{m_1}$ y $Z_n = X_n \pmod{m_2}$, $\forall n \geq 0$.

Demostremos primero que

$$X_n = X_k \Leftrightarrow Y_n = Y_k \text{ y } Z_n = Z_k .$$

La implicación \Rightarrow es evidente.

Veamos \Leftarrow :

$$\begin{aligned} \bullet Y_n = Y_k &\Rightarrow X_n \bmod m_1 = X_k \bmod m_1 \Rightarrow \begin{cases} X_n = r_1 + m_1 q_1 \\ X_k = r_1 + m_1 q'_1 \end{cases} \\ \bullet Z_n = Z_k &\Rightarrow X_n \bmod m_2 = X_k \bmod m_2 \Rightarrow \begin{cases} X_n = r_2 + m_2 q_2 \\ X_k = r_2 + m_2 q'_2 \end{cases} \end{aligned}$$

Luego

$$\begin{cases} X_n - X_k = m_1(q_1 - q'_1) \\ X_n - X_k = m_2(q_2 - q'_2) \end{cases} \Rightarrow \begin{cases} X_n - X_k = m_1 \dot{m} \\ X_n - X_k = m_2 \dot{m} \end{cases} \Rightarrow X_n - X_k = \dot{m} ,$$

y esto es imposible ya que X_n y X_k son restos de dividir por m , luego estaríamos ante una contradicción, a no ser que $X_n - X_k = 0 \Rightarrow X_n = X_k$, que es lo que queríamos demostrar.

Volviendo a la demostración del lema, sean λ_1 la longitud del periodo de la secuencia $\{Y_n\}$ y λ_2 la de la secuencia $\{Z_n\}$. Sea $\tilde{\lambda} = \text{mcm}(\lambda_1, \lambda_2)$. Veamos que $\lambda = \tilde{\lambda}$:

Como $X_n = X_{n+\lambda}$, para un n suficientemente grande, entonces

$$\begin{cases} Y_n = Y_{n+\lambda} \Rightarrow \lambda \text{ es múltiplo de } \lambda_1 \\ Z_n = Z_{n+\lambda} \Rightarrow \lambda \text{ es múltiplo de } \lambda_2 \end{cases} ,$$

de donde

$$\lambda \geq \tilde{\lambda} . \quad (2.4)$$

Por otro lado, como $\tilde{\lambda}$ es múltiplo de $\lambda_1 \Rightarrow Y_n = Y_{n+\tilde{\lambda}}$ y por ser $\tilde{\lambda}$ múltiplo de $\lambda_2 \Rightarrow Z_n = Z_{n+\tilde{\lambda}}$, para n suficientemente grande. Por lo que $X_n = X_{n+\tilde{\lambda}}$, de donde

$$\lambda \leq \tilde{\lambda} . \quad (2.5)$$

Uniendo las desigualdades (2.4) y (2.5) tenemos la igualdad buscada. \square

Estos dos lemas nos dan las herramientas necesarias para demostrar el teorema que nos indica qué elección de parámetros es la adecuada para que un G.C.L. tenga longitud de periodo máxima:

Teorema 2.1 *La secuencia de números obtenida mediante un G.C.L. de parámetros (X_0, a, c, m) tiene periodo de longitud máxima si, y sólo si,*

1. c y m son primos relativos.
2. $a \equiv 1 \pmod{p}$, para todo p factor primo de m .
3. $a \equiv 1 \pmod{4}$, si m es múltiplo de 4.

Demostración: Por el lema 2.2, basta probar el teorema cuando m es potencia de un número primo, pues al ser $\lambda \leq m$, la secuencia alcanzará la longitud de periodo máxima cuando $\lambda = m$, es decir, si

$$m = p_1^{e_1} \cdot \dots \cdot p_t^{e_t} = \lambda = \text{mcm}(\lambda_1, \dots, \lambda_t) \leq \lambda_1 \cdot \dots \cdot \lambda_t \leq p_1^{e_1} \cdot \dots \cdot p_t^{e_t} ,$$

y esto sólo puede ser cierto si, y sólo si, $\lambda_j = p_j^{e_j}$, $1 \leq j \leq t$. Por tanto supondremos $m = p^e$ donde p es primo y $e \in \mathbb{N}$.

Como ya comentamos al comienzo del capítulo, el caso $a = 1$ es trivial, por lo que tomaremos $a > 1$.

El periodo será de longitud m si, y sólo si, todos los enteros $0 \leq x < m$ aparecen en el periodo (ya que ningún valor puede repetirse); por lo tanto, el periodo es de longitud m si, y sólo si, la longitud del periodo de la secuencia para $X_0 = 0$ es m .

Mediante la proposición 2.1, podemos expresar el n -ésimo término de la secuencia en función de X_0 como:

$$X_{n+k} = \left(a^k X_{kn} + \frac{a^k - 1}{a - 1} \cdot c \right) \pmod{m} \Rightarrow X_n = \left(a^n X_0 + \frac{a^n - 1}{a - 1} \cdot c \right) \pmod{m} .$$

Para $X_0 = 0$, tendremos

$$X_n = \left(\frac{a^n - 1}{a - 1} \cdot c \right) \pmod{m} \tag{2.6}$$

Si c no es primo relativo de m , $X_n \neq 1$ para todo n . Veamos esto:

Supongamos que c y m no son primos relativos y al ser m potencia de un primo, entonces $(c, m) = p^k \neq 1$ y supongamos, para llegar a una contradicción que $X_n = 1$ para algún n , entonces, aplicando la igualdad (2.6),

$$\begin{aligned} \frac{a^n - 1}{a - 1} \cdot c = X_n + q \cdot m &\Rightarrow \frac{a^n - 1}{a - 1} \cdot c = 1 + q \cdot m \Rightarrow \\ \frac{a^n - 1}{a - 1} \cdot c - q \cdot m = 1 &\Rightarrow \left(\frac{a^n - 1}{a - 1} \cdot c' - q \cdot m' \right) \cdot p^k = 1 \end{aligned}$$

y llegaríamos a una contradicción, por lo que $X_n \neq 1$. Queda así demostrada la necesidad de la primera condición.

Falta pues, para finalizar la demostración del teorema comprobar que

$$\lambda = m = p^e \Leftrightarrow \begin{cases} a \equiv 1 \pmod{p}, & \text{cuando } p > 2 \\ a \equiv 1 \pmod{4}, & \text{cuando } p=2 \end{cases} .$$

Como $(c, m) = 1$, aplicando la igualdad (2.6),

$$X_n = \frac{a^n - 1}{a - 1} \pmod{m} .$$

Por otro lado, como λ es la longitud del periodo, entonces λ es el menor entero positivo que cumple que

$$X_\lambda = X_0 = 0 \Rightarrow \frac{a^\lambda - 1}{a - 1} \pmod{m} = 0 \Rightarrow \frac{a^\lambda - 1}{a - 1} \equiv 0 \pmod{m} . \quad (2.7)$$

Veamos primero la implicación a la derecha:

$\lambda = p^e$ y supongamos, para llegar a una contradicción, que $a \not\equiv 1 \pmod{p}$, entonces

$$\frac{a^n - 1}{a - 1} \equiv 0 \pmod{p^e} \Leftrightarrow a^n - 1 \equiv 0 \pmod{p^e}$$

Luego

$$\begin{aligned} \frac{a^n - 1}{a - 1} \equiv 0 \pmod{m} &\Rightarrow a^\lambda - 1 \equiv 0 \pmod{\lambda} \Rightarrow \\ a^{p^e} - 1 \equiv 0 \pmod{p^e} &\Rightarrow a^{p^e} \equiv 1 \pmod{p} \end{aligned}$$

Pero por el Teorema de Fermat sabemos que si p es un número primo, entonces $a^p \equiv a \pmod{p}$, de esto y aplicando inducción sobre e obtenemos que $a^{p^e} \equiv a^e \pmod{p}$, llegando así a una contradicción que proviene de suponer que $a \not\equiv 1 \pmod{p}$ cuando $p > 2$.

Si $p = 2$, tenemos que demostrar que $a \equiv 1 \pmod{4}$. Las posibilidades que hay es que a sea congruente con 0, con 1, con 2 o con 3.

Si $a \equiv 2 \pmod{4}$, entonces existe algún entero q tal que $a = 4q + 2$, con lo que obtenemos que

$$a^2 = 16q^2 + 4 + 8q = 2 \pmod{4} \Rightarrow a^2 \equiv 0 \pmod{4} ,$$

es decir,

$$a^{2^1} - 1 \not\equiv 0 \pmod{2^1} \Rightarrow \frac{a^{2^1} - 1}{a - 1} \not\equiv 0 \pmod{2^1} .$$

Aplicando este razonamiento para 2^2 , y dado que a es múltiplo de 2 (por ejemplo $a = 2k$),

$$a^{2^2} = (2k)^4 = 2^2(4k^4) \Rightarrow a^{2^2} \equiv 0 \pmod{2^2} \Rightarrow a^{2^2} - 1 \not\equiv 0 \pmod{2^2} .$$

De igual forma, llegamos a

$$a^{2^e} - 1 \not\equiv 0 \pmod{2^e} ,$$

y por lo tanto

$$\frac{a^{2^e} - 1}{a - 1} \not\equiv 0 \pmod{2^e} ,$$

lo que entra en contradicción con la expresión (2.7).

Si $a \equiv 0 \pmod{4}$ entonces quiere decir que a es múltiplo de 4 y, además, múltiplo de 2. Por lo tanto, llegamos a una contradicción siguiendo el mismo razonamiento que para $a \equiv 2 \pmod{4}$.

Por último, veamos qué ocurre si $a \equiv 3 \pmod{4}$. Por un lado, vemos que se cumple que $a^2 \equiv 1 \pmod{8}$, $a^4 \equiv 1 \pmod{16}$, $a^8 \equiv 1 \pmod{32}$, etc. Es decir, en general

$$a^{2^{e-1}} - 1 \equiv 0 \pmod{2^{e+1}} .$$

Consecuencia de esto es que

$$\frac{a^{2^{e-1}} - 1}{2} \equiv 0 \pmod{\frac{2^{e+1}}{2}} . \quad (2.8)$$

Por otro lado,

$$a \equiv 3 \pmod{4} \Rightarrow a = 4q + 3 \Rightarrow a - 1 = 4q + 2 \Rightarrow a - 1 = 2(2q + 1) ,$$

es decir, $a - 1$ es el doble de un número impar. Este hecho, junto con (2.8) implica que

$$\frac{a^{2^{e-1}} - 1}{a - 1} \equiv 0 \pmod{2^e} . \quad (2.9)$$

Pero como partimos de que $\lambda = 2^e$, y $2^{e-1} < \lambda$, llegamos a una contradicción con (2.7).

Por lo tanto, la única posibilidad es que $a \equiv 1 \pmod{4}$.

Veamos la implicación a la izquierda:

Como $a \equiv 1 \pmod{p}$ y aplicando sucesivamente el lema (2.1) llegamos a

$$a^{p^g} \equiv 1 \pmod{p^{f+g}}, \quad a^{p^g} \not\equiv 1 \pmod{p^{f+g+1}} \quad \forall g \geq 0$$

de donde

$$\begin{aligned} \frac{a^{p^g} - 1}{a - 1} &\equiv 0 \pmod{p^g} , \\ \frac{a^{p^g} - 1}{a - 1} &\not\equiv 0 \pmod{p^{g+1}} . \end{aligned} \quad (2.10)$$

En particular, $\frac{a^{p^e} - 1}{a - 1} \equiv 0 \pmod{p^e}$, es decir, $X_{p^e} = 0$, y por ser λ la longitud del periodo, p^e debe ser múltiplo de λ . Así $\lambda = p^g$ para algún g , y por la ecuación (2.10) $\lambda = p^e$, completando así la demostración. \square

2.2.1.2 Generador Congruencial Multiplicativo (G.C.M.)

Es un caso particular de los lineales cuando $c = 0$.

Vienen dados por la expresión

$$X_{n+1} = (aX_n) \pmod{m} . \quad (2.11)$$

Como $X_0 \neq 0$, el teorema principal para los G.C.L. nos muestra que no puede alcanzarse el periodo máximo. A cambio este método es más rápido al tener menos operaciones.

Buscaremos condiciones bajo las cuales este generador alcance periodos con la mayor longitud posible.

Por el lema (2.2) podemos considerar $m = p^e$, obteniendo la expresión

$$X_n = (a^n X_0) \pmod{p^e} .$$

Si a es múltiplo de p , el periodo será de longitud 1, por lo que a y p deberán ser primos relativos. Veamos algunas definiciones necesarias para encontrar las condiciones de periodo máximo.

Definición 2.3 Sean $a, m \in \mathbb{Z}$ primos relativos. Se define el orden de a módulo m , y se denota por $\lambda(a)$ al menor entero que verifica,

$$a^{\lambda(a)} \equiv 1 \pmod{m} .$$

Definición 2.4 Sea $m \in \mathbb{Z}$. Se define el orden módulo m , y se denota por $\lambda(m)$, como

$$\lambda(m) = \max_{a < m} \lambda(a) .$$

Definición 2.5 Sean $a, m \in \mathbb{Z}$. Se dice que a es raíz primitiva de m si y sólo si $\lambda(a) = \lambda(m)$.

Ya podemos enunciar el teorema que caracteriza los generadores congruenciales multiplicativos de periodo máximo.

Teorema 2.2 *El periodo máximo para un G.C.M. con módulo m es de longitud $\lambda(m)$, y se alcanza si y sólo si,*

1. X_0 es primo relativo con m .
2. a es raíz primitiva de m .

Si m es primo, podemos obtener una secuencia con periodo de longitud $m - 1$, por lo que en la práctica podremos obtener periodos tan largos como queramos.

La cuestión ahora es cómo encontrar raíces primitivas de m , la respuesta a esta pregunta nos la da el siguiente teorema:

Teorema 2.3 *a es una raíz primitiva módulo p^e si y sólo si se verifica una de las siguientes condiciones:*

1. Si $p^e = 2$ y a es primo;
2. Si $p^e = 4$ y $a \bmod 4 = 3$;
3. Si $p^e = 8$ y $a \bmod 8 = 3, 5, 7$;
4. Si $p = 2$, $e \geq 4$ y $a \bmod 8 = 3, 5$;
5. Si p es primo, $e = 1$, $a \not\equiv 0 \pmod{p}$ y $a^{\frac{p-1}{q}} \not\equiv 1 \pmod{p}$ siendo q cualquier primo divisor de $(p - 1)$;
6. Si p es primo, $e > 1$, $a \not\equiv 0 \pmod{p}$, $a^{\frac{p-1}{q}} \not\equiv 1 \pmod{p}$ siendo q cualquier primo divisor de $(p - 1)$ y $a^{p-1} \not\equiv 1 \pmod{p^2}$.

2.2.1.3 Otros generadores congruenciales

Para concluir con los generadores congruenciales, enunciaremos sin más algunas otras posibilidades menos utilizadas que los G.C.L. y G.C.M.

Generador congruencial cuadrático

Viene definido por la expresión:

$$X_{n+1} = (dX_n^2 + aX_n + c) \pmod{m} .$$

El periodo máximo es m .

Generador congruencial lineal generalizado

Viene definido por la expresión:

$$X_n = (a_1X_{n-1} + a_2X_{n-2} + \cdots + a_kX_{n-k}) \pmod{m} .$$

El periodo máximo es de longitud $m^k - 1$ si m es primo.

Generador aditivo de Fibonacci

Se define como

$$X_{n+1} = (X_n + X_{n-1}) \pmod{m} .$$

Generador aditivo de Green

$$X_{n+1} = (X_n + X_{n-k}) \pmod{m} .$$

Generador aditivo de Mitchell-Moore

$$X_n = (X_{n-24} + X_{n-55}) \pmod{m}$$

Para:

- m par y grande.
- X_0, \dots, X_{54} enteros menores que m y no todos pares.

Bajo estas condiciones el ciclo máximo es de $2^{55} - 1$.

2.3 Métodos de mejora de la aleatoriedad

En esta sección estudiaremos la posibilidad de refinar las secuencias obtenidas por los generadores estudiados anteriormente.

2.3.1 Método aditivo de dos series

Consiste en, dadas dos secuencias de números aleatorios, crear otra nueva sumando estas dos.

Sean $\{X_n\}$ e $\{Y_n\}$ dichas secuencias con períodos λ_1 y λ_2 respectivamente, entonces podemos generar $\{Z_n\}$ mediante

$$Z_n = (X_n + Y_n) \pmod{m} .$$

Con esto conseguimos aumentar la longitud del periodo, pues si λ_1 y λ_2 son primos relativos, entonces,

$$\lambda_z = \lambda_1 \cdot \lambda_2 .$$

2.3.2 Algoritmo de mezcla de dos series

Consiste en reordenar los términos de una sucesión usando para ello una sucesión auxiliar. Sea $\{X_n\}$ módulo m , e $\{Y_n\}$ módulo m' . Se reordena X_n atendiendo a Y_n . Se toma un vector con los k primeros elementos de la sucesión, siendo $k \simeq 100$.

$v(0)$	\cdots	$v(k-1)$
\uparrow		\uparrow
X_0	\cdots	X_{k-1}

Dado un número y de la secuencia $\{Y_n\}$, se calcula una posición del vector v :

$$h = \left\lfloor \frac{k \cdot y}{m'} \right\rfloor ,$$

con $0 \leq h \leq k - 1$. Entonces se saca el valor almacenado en $v(h)$ como siguiente número aleatorio, y se reemplaza en el vector por el siguiente elemento de la sucesión $\{X_n\}$. El algoritmo es el siguiente:

ALGORITMO MEZCLA 1

1. Desde $j := 0$ hasta $k - 1$
 - $v(j) := X_j$
2. $i := 0$
3. $x := X_{k+i}$
4. $y := Y_i$
5. $h := \left\lfloor \frac{k \cdot y}{m'} \right\rfloor$
6. Dar $v(h)$ como número de la nueva secuencia.
7. $v(h) := x$
8. $i := i + 1$
9. Mientras queden números por generar
 - Ir al paso 3.

2.3.3 Algoritmo de mezcla con una sola serie

La idea es utilizar la propia sucesión de partida para realizar la reordenación. En un primer momento se usa el X_k para sortear la posición dentro del vector v , y a partir de ahí se usa el mismo $v(h)$ que se dé como salida para el sorteo de la siguiente posición.

ALGORITMO MEZCLA 2

1. Desde $j := 0$ hasta $k - 1$
 - $v(j) := X_j$
2. $j := k$
3. $y := X_k$
4. $h := \left\lfloor \frac{k \cdot y}{m} \right\rfloor$
5. $y := v(h)$
6. Dar $v(h)$ como número de la nueva secuencia.
7. $j := j + 1$
8. $v(h) := X_j$
9. Mientras queden números por generar
 - Ir al paso 4.

Capítulo 3

Comprobación de la aleatoriedad

En este capítulo revisaremos algunos contrastes empíricos para analizar la aleatoriedad de una secuencia de datos, procedente de un generador de los vistos anteriormente. En este sentido, las pruebas que diseñemos han de contrastar tanto la uniformidad como la independencia de los datos. Usaremos en todos los casos contrastes bilaterales, en el sentido de que rechazaremos la hipótesis de aleatoriedad tanto si los datos se alejan claramente de la aleatoriedad, en cuanto a uniformidad e independencia, como si se acercan demasiado a ella. Es decir, rechazaremos también una secuencia por ser “demasiado aleatoria”, pues este hecho puede producir sospechas de que los datos han sido generados intencionadamente para producir esos resultados.

3.1 Test χ^2

Este contraste sirve para analizar la aleatoriedad de los datos en cuanto a frecuencias. Se trata de contrastar la hipótesis de que los datos proceden de una distribución uniforme en el intervalo $[0, 1)$.

El test χ^2 está orientado a datos categóricos, por lo que para aplicarlo al caso de números aleatorios hemos de agrupar los datos en categorías.

El punto de partida será una secuencia de n números X_1, \dots, X_n , $X_i \in [0, 1)$, $i = 1, \dots, n$, cuya aleatoriedad queremos contrastar.

El proceso de análisis lo podemos desglosar en los siguientes pasos:

1. Dividir el intervalo $[0, 1)$ en d partes iguales, que llamaremos *clases*.
2. Contabilizar cuántos números caen en cada clase. A la cantidad de números que caen en una clase $i = 1, \dots, d$ la llamaremos *frecuencia observada de la clase i* , y la denotaremos por θ_i . (Obsérvese que $\sum_{i=1}^d \theta_i = n$).
3. Comparamos las frecuencias observadas con las *frecuencias esperadas E_i* , es decir, con la cantidad de números que esperaríamos encontrar en la clase i si la secuencia fuera realmente aleatoria.

La frecuencia esperada E_i se puede calcular como

$$E_i = np_i \quad , \quad (3.1)$$

donde p_i es la probabilidad de que un número elegido al azar entre 0 y 1 caiga en la clase i . Esta probabilidad es $p_i = 1/d$, dado que todas las clases son de igual amplitud, con lo que

$$E_i = n \frac{1}{d} = \frac{n}{d} \quad . \quad (3.2)$$

A la hora de implementar este test, es útil detectar de forma sencilla la clase en que cae cada número de la secuencia. Si llamamos C_i a la clase en la que cae el número X_i , se cumple que

$$C_i = \lfloor d \times x_i \rfloor + 1 \quad , \quad (3.3)$$

donde $\lfloor \cdot \rfloor$ es la función parte entera.

Una vez llegados a este punto, podemos medir el ajuste de los datos a una distribución $\mathcal{U}[0, 1)$ usando el estadístico

$$\chi^2 = \sum_{i=1}^d \frac{(\theta_i - E_i)^2}{E_i} \quad , \quad (3.4)$$

que se distribuye asintóticamente según una distribución $\chi^2(d-1)$. Una condición **imprescindible** para poder aplicar el test χ^2 es que las frecuencias esperadas en cada clase sean siempre mayores que 5, es decir,

$$E_i = \frac{n}{d} > 5 ,$$

de manera que elegiremos el número de clases d teniendo en cuenta esta restricción.

Podemos construir un contraste para la hipótesis nula H_0 : *los datos proceden de $\mathcal{U}[0, 1)$* frente a H_1 : *los datos no proceden de $\mathcal{U}[0, 1)$* . Fijando un nivel de significación $\alpha = 0.05$, la región crítica es

$$\{\chi^2 \leq \chi_{d-1, \alpha/2}^2\} \cup \{\chi^2 \geq \chi_{d-1, 1-\alpha/2}^2\} =$$

$$\{\chi^2 \leq \chi_{d-1, 0.025}^2\} \cup \{\chi^2 \geq \chi_{d-1, 0.975}^2\} .$$

A la hora de calcular el valor del estadístico, podemos aprovechar el hecho de que los E_i son constantes para obtener una fórmula abreviada:

$$\begin{aligned} \chi^2 &= \sum_{i=1}^d \frac{(\theta_i - E_i)^2}{E_i} = \sum_{i=1}^d \frac{(\theta_i - (n/d))^2}{n/d} \\ &= \frac{d}{n} \sum_{i=1}^d \left(\theta_i - \frac{n}{d}\right)^2 \\ &= \frac{d}{n} \sum_{i=1}^d \left(\theta_i^2 + \frac{n^2}{d^2} - 2\theta_i \frac{n}{d}\right) \\ &= \frac{d}{n} \left(\sum_{i=1}^d \theta_i^2 + \sum_{i=1}^d \frac{n^2}{d^2} - 2\frac{n}{d} \sum_{i=1}^d \theta_i \right) \\ &= \frac{d}{n} \sum_{i=1}^d \theta_i^2 + \frac{d}{n} \cdot \frac{n^2}{d} - 2\frac{d}{n} \cdot \frac{n^2}{d} \\ &= \frac{d}{n} \sum_{i=1}^d \theta_i^2 - n . \end{aligned}$$

Ejemplo 3.1 Supongamos que deseamos analizar la aleatoriedad de una secuencia de 100 números aplicando el test χ^2 , y que hemos dividido el intervalo $[0, 1)$ en 5 clases (obsérvese que $n/d = 100/5 = 20 > 5$), obteniendo las siguientes frecuencias observadas:

i	1	2	3	4	5
$\theta(i)$	25	21	15	19	20

Con esto podemos calcular el valor del estadístico

$$\chi^2 = \frac{5}{20}(19^2 + 21^2 + 20^2 + 18^2 + 22^2) - 100 = 0.5$$

La región crítica es

$$\{\chi^2 \leq 0.48\} \cup \{\chi^2 \geq 11.14\} ,$$

con lo que NO rechazaríamos la hipótesis nula, al encontrarse el valor del estadístico fuera de ella.

3.2 Test de series

Con este contraste se pretende analizar la aleatoriedad de los datos en cuanto a la independencia. La idea fundamental del test consiste en agrupar los datos en parejas y comprobar si existe cierta tendencia no aleatoria en dichas parejas: por ejemplo, que a números pequeños tiendan a suceder números grandes u otras situaciones de ese carácter.

Partiremos de una secuencia de números X_1, \dots, X_n . Consideraremos, sin pérdida de generalidad, que la cantidad de números es par (si no lo es, desechamos el último número). De esta forma, los n datos de partida dan lugar a $m = n/2$ parejas $(X_1, X_2), \dots, (X_{n-1}, X_n)$.

Cada pareja podemos verla como un punto de la región correspondiente al hiper-cubo unidad en \mathbb{R}^2 , de manera que si los números fueran aleatorios, las parejas se distribuirían uniformemente en dicha región.

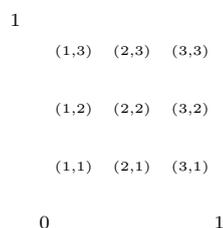


Figura 3.1: División de la región $[0, 1) \times [0, 1)$ para $d = 3$.

Procedemos de forma similar al test χ^2 explicado anteriormente, dividiendo la región en zonas iguales. Esto lo hacemos dividiendo el intervalo $[0, 1)$ en d clases iguales, con lo que obtenemos una división de la región $[0, 1) \times [0, 1)$ en $d \times d = d^2$ clases o celdas iguales (ver figura 3.1).

A continuación comprobamos en qué clase cae cada pareja (X_i, X_{i+1}) , para lo cual basta con aplicar la fórmula $(\lfloor X_i \times d \rfloor + 1, \lfloor X_{i+1} \times d \rfloor + 1)$. Por ejemplo, si tenemos $d = 3$, la pareja $(0.1, 0.5)$ se corresponde con la casilla

$$(\lfloor 3 \times 0.1 \rfloor + 1, \lfloor 3 \times 0.5 \rfloor + 1) = (\lfloor 0.3 \rfloor + 1, \lfloor 1.5 \rfloor + 1) = (1, 2) .$$

A continuación calculamos las frecuencias observadas en cada celda, es decir, la cantidad de parejas que caen en cada celda, que denotaremos θ_{ij} , $1 \leq i, j \leq d$, y las compararemos con las frecuencias esperadas E_{ij} , $1 \leq i, j \leq d$.

Obsérvese que bajo la hipótesis nula de que los números son aleatorios, estas frecuencias deben ser iguales para todas las clases; como tenemos d^2 clases de igual tamaño, la probabilidad de que una pareja al azar caiga en la casilla (i, j) es $p_{ij} = 1/d^2$, y por tanto,

$$E_{ij} = \frac{m}{d^2} .$$

Como estadístico de contraste usaremos

$$\chi^2 = \sum_{i=1}^d \sum_{j=1}^d \frac{(\theta_{ij} - E_{ij})^2}{E_{ij}} , \tag{3.5}$$

que sigue asintóticamente una distribución $\chi^2(d^2 - 1)$.

En este caso también nos encontramos con la restricción de que las frecuencias esperadas han de ser mayores que 5, lo que hay que tener en cuenta al elegir el número de divisiones d .

Podemos construir un contraste para la hipótesis nula H_0 : *los datos son independientes y $\mathcal{U}[0, 1)$* frente a H_1 : *los datos no son independientes y $\mathcal{U}[0, 1)$* . Fijando un nivel de significación $\alpha = 0.05$, la región crítica es

$$\{\chi^2 \leq \chi_{d^2-1, \alpha/2}^2\} \cup \{\chi^2 \geq \chi_{d^2-1, 1-\alpha/2}^2\} = \{\chi^2 \leq \chi_{d^2-1, 0.025}^2\} \cup \{\chi^2 \geq \chi_{d^2-1, 0.975}^2\} .$$

3.3 Test de rachas

Este contraste trata de detectar la presencia de rachas crecientes o decrecientes. Consideraremos aquí el caso de rachas crecientes.

De nuevo partimos de una secuencia de números X_1, \dots, X_n . Llamaremos *racha de longitud r* a una secuencia de la forma

$$X_1 < X_2 < X_3 < \dots < X_r > X_{r+1} .$$

Considerando las longitudes de las rachas presentes en la secuencia de partida, construimos una nueva variable que mide dichas longitudes, y sobre la cual desarrollaremos un contraste basado en la χ^2 . Dado que a priori no sabemos cuál es la mayor longitud de racha que vamos a encontrar, estableceremos una longitud máxima t^* , considerando todas las rachas que pudieran aparecer de longitud mayor que t^* como si fueran de longitud igual a t^* . La mínima racha será aquella en la que el elemento que estamos explorando sea mayor que el siguiente, en cuyo caso la racha tendrá longitud 1, con lo que el conjunto de posibles valores de la variable *longitud* será

$$\{1, 2, \dots, t^*\} .$$

Para evitar comenzar siempre las rachas con valores pequeños, una vez terminemos de contabilizar una racha saltaremos un número antes de empezar a contar la siguiente, ya que de otra forma siempre empezaríamos a contarlas desde el elemento que ha roto la secuencia creciente, y que por tanto tendrá cierta tendencia a ser pequeño.

Una vez encontradas las rachas, calculamos cuántas han aparecido de cada una de las posibles longitudes, es decir, contabilizamos las frecuencias observadas de cada longitud de racha, que denotaremos de nuevo por θ_i , $i = 1, \dots, t^*$. Estas frecuencias observadas las enfrentaremos a las frecuencias esperadas E_i , $i = 1, \dots, t^*$ mediante el estadístico χ^2 .

Para calcular las frecuencias esperadas, necesitamos conocer la probabilidad de que una racha tenga longitud i bajo la hipótesis nula de que los números son aleatorios. Concretamente, esta probabilidad se calcula como:

$$p_i = \frac{i}{(i+1)!} \quad i = 1, \dots, t^* - 1 ,$$

$$p_{t^*} = \frac{1}{t^*!} ,$$

las cuales se pueden obtener simplemente computando casos favorables y casos posibles.

Con esto, tenemos que

$$E_i = \frac{n \cdot i}{(i+1)!} \quad i = 1, \dots, t^* - 1 ,$$

$$E_{t^*} = \frac{n}{t^*!} .$$

Con estos valores calculamos el valor del estadístico

$$\chi^2 = \sum_{i=1}^{t^*} \frac{(\theta_i - E_i)^2}{E_i} , \quad (3.6)$$

cuya distribución es asintóticamente $\chi^2(t^* - 1)$.

Podemos construir un contraste para la hipótesis nula H_0 : *los datos son independientes y $\mathcal{U}[0, 1)$* frente a H_1 : *los datos no son independientes y $\mathcal{U}[0, 1)$* . Fijando un nivel de significación $\alpha = 0.05$, la región crítica es

$$\{\chi^2 \leq \chi_{t^*-1, \alpha/2}^2\} \cup \{\chi^2 \geq \chi_{t^*-1, 1-\alpha/2}^2\} =$$

$$\{\chi^2 \leq \chi_{t^*-1, 0.025}^2\} \cup \{\chi^2 \geq \chi_{t^*-1, 0.975}^2\} .$$

3.4 Test de huecos o de distancias

Este contraste analiza la tendencia de los datos a concentrarse dentro de cierto subintervalo del $[0, 1)$.

De nuevo partimos de una secuencia de n números X_1, \dots, X_n . Consideramos un intervalo $[\alpha, \beta)$ con $0 \leq \alpha < \beta \leq 1$.

Llamaremos *hueco* a cualquier serie de números en la secuencia de partida, que no están dentro del intervalo $[\alpha, \beta)$ y que están comprendidos entre dos números que sí lo están. A la cantidad de números en ese hueco que NO pertenecen a $[\alpha, \beta)$ la llamaremos *longitud del hueco*.

Ejemplo 3.2 *Sea la secuencia $\{0.7, 0.3, 0.1, 0.2, 0.8, 0.75\}$. Si consideramos el intervalo $[0.6, 0.9)$, los números $0.7, 0.3, 0.1, 0.2, 0.8$ forman un hueco de longitud 3, mientras que los números $0.8, 0.75$ forman un hueco de longitud 0.*

Por tanto, las posibles longitudes de los huecos oscilarán entre 0 y un valor máximo, establecido por nosotros, t^* . Si aparece algún hueco de longitud mayor que t^* , procederemos igual que en el caso del test de rachas: considerándolo como uno de longitud t^* .

Contabilizamos las frecuencias observadas θ_i , $i = 0, \dots, t^*$ y las frecuencias esperadas $E_i = n \cdot p_i$, $i = 0, \dots, t^*$. La probabilidad de obtener un hueco de longitud i viene dada por:

longitud	0	1	2	...	t^*
p_i	p	$(1-p)p$	$(1-p)^2p$...	$(1-p)^{t^*}$

donde $p = \beta - \alpha$ es la probabilidad de que un número al azar caiga dentro del intervalo $[\alpha, \beta)$. Proponemos como ejercicio para el lector comprobar que las probabilidades de la tabla anterior son correctas.

Una vez obtenidas las frecuencias, calculamos el valor del estadístico

$$\chi^2 = \sum_{i=0}^{t^*} \frac{(\theta_i - E_i)^2}{E_i}, \quad (3.7)$$

cuya distribución es asintóticamente $\chi^2(t^*)$.

Podemos construir un contraste para la hipótesis nula H_0 : *los datos son independientes y $\mathcal{U}[0, 1)$* frente a H_1 : *los datos no son independientes y $\mathcal{U}[0, 1)$* . Fijando un nivel de significación $\alpha = 0.05$, la región crítica es

$$\{\chi^2 \leq \chi_{t^*, \alpha/2}^2\} \cup \{\chi^2 \geq \chi_{t^*, 1-\alpha/2}^2\} =$$

$$\{\chi^2 \leq \chi_{t^*, 0.025}^2\} \cup \{\chi^2 \geq \chi_{t^*, 0.975}^2\} .$$

3.5 Test de Kolmogorov-Smirnov

Este contraste mide el ajuste de la distribución empírica de los datos a la distribución teórica, es decir, la $\mathcal{U}[0, 1)$.

Partimos de una secuencia de datos X_1, \dots, X_n . En primer lugar, ordenamos de menor a mayor la secuencia anterior. Sea $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ la secuencia ordenada.

La *función de distribución empírica* correspondiente a los datos observados viene dada por

$$G(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x)}(X_i) ,$$

donde

$$I_{(-\infty, x)}(X) = \begin{cases} 1 & \text{si } -\infty < X \leq x \\ 0 & \text{en otro caso.} \end{cases}$$

Es decir, $G(x)$ mide la proporción de valores en la muestra menores o iguales que x .

Ejemplo 3.3 Sea la muestra $\{0.2, 0.5, 0.7, 0.9\}$. La figura 3.2 muestra la función de distribución empírica correspondiente a esta muestra en contraste con la función de distribución de la $\mathcal{U}[0, 1)$.

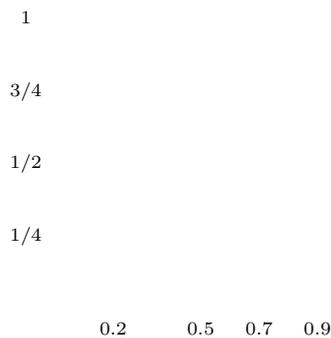


Figura 3.2: Distribuciones teórica y empírica.

Denotemos por $G(x)$ a la función de distribución empírica y $F(x)$ a la función de distribución de la $\mathcal{U}[0, 1)$, y sea $G^+(X_i) = \lim_{x \rightarrow X_i^+} G(x)$.

Una posibilidad para medir la distancia entre la distribución empírica y la distribución teórica es mediante el estadístico

$$\begin{aligned} D_n^+ &= \sup_{-\infty < x < \infty} (G(x) - F(x)) \\ &= \max_{1 \leq i \leq n} \{G^+(X_i) - F(X_i)\} \\ &= \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(X_i) \right\} \end{aligned}$$

que mide la mayor diferencia favorable a la distribución empírica.

Para muestras grandes, podemos construir un test basado en D^+ aprovechando el hecho de que

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n^+ \leq x) = 1 - e^{-2x^2}, \quad x > 0 .$$

Con esto, tenemos un contraste a nivel de significación $\alpha = 0.05$ para la hipótesis nula H_0 : los datos proceden de una distribución uniforme en $[0, 1)$, frente a H_1 : los datos no proceden de una distribución uniforme en $[0, 1)$ con región crítica

$$\{\sqrt{n}D_n^+ \leq h_{\alpha/2}\} \cup \{\sqrt{n}D_n^+ \geq h_{1-\alpha/2}\} =$$

$$\{\sqrt{n}D_n^+ \leq h_{0.025}\} \cup \{\sqrt{n}D_n^+ \geq h_{0.975}\} =$$

$$\{\sqrt{n}D_n^+ \leq 0.1125\} \cup \{\sqrt{n}D_n^+ \geq 1.3581\} ,$$

donde h_α es el cuantil de orden α de la distribución $H(x) = P(\sqrt{n}D_n^+ \leq x)$.

Capítulo 4

Generación de muestras

Disponiendo de un mecanismo que proporcione números aleatorios, es posible generar automáticamente una muestra a partir de una distribución de probabilidad determinada. La posibilidad de contar con este tipo de muestras *artificiales* es de gran utilidad. Por ejemplo, pueden servirnos para contrastar nuevos procedimientos estadísticos, para la aplicación de técnicas de remuestreo (posibilidad que analizaremos en posteriores capítulos) o para simular valores de las variables que intervienen en un modelo de simulación. A partir de ahora, entenderemos por *simular una variable* el proceso de obtener valores para dicha variable.

A los métodos que propongamos para simular variables, les exigiremos las siguientes características:

1. **Exactitud.** La muestra obtenida debe ajustarse a la distribución de la variable que estamos simulando.
2. **Eficiencia.** Hay que intentar minimizar el número de operaciones necesarias para generar los valores de la variable.
3. **Robustez.** Los métodos diseñados han de servir para cualquier valor de los parámetros de la distribución de la variable a simular.

4.1 Método de inversión

El primer método que veremos se basa en usar la función de distribución de la variable para la cual queremos obtener valores.

Teorema 4.1 *Sea X una v.a. con función de distribución F_X . Si $U \rightsquigarrow \mathcal{U}(0,1)$, entonces la variable $Y = F_X^{-1}(U)$ tiene la misma distribución que X .*

Demostración: Podemos definir

$$F_X^{-1}(y) = \inf \{x \in \mathbb{R} \mid F_X(x) \geq y\} \quad 0 \leq y \leq 1 .$$

Entonces,

$$F_Y(x) = P\{Y \leq x\} = P\{F_X^{-1}(U) \leq x\} = P\{U \leq F_X(x)\} = F_X(x) .$$

Por tanto, X e Y tienen la misma distribución. □

El teorema anterior indica una posibilidad para generar una muestra para la variable X . Basta con generar números aleatorios y calcular sus inversos mediante F_X , obteniendo una muestra de variables independientes con la misma distribución que X . Este proceso se muestra en la figura 4.1.

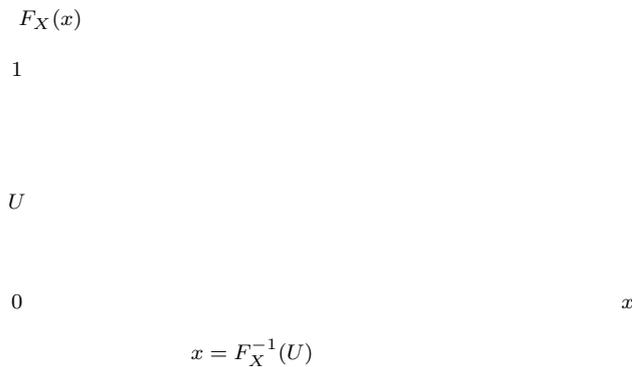


Figura 4.1: Método de inversión.

Ejemplo 4.1 Supongamos que deseamos generar una muestra de tamaño 3 de una v.a. X con distribución $\mathcal{U}(2, 4)$. La función de distribución de X es

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 2 \\ \frac{x-2}{2} & \text{si } 2 < x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

Dado $0 < u < 1$, la inversa de la función de distribución viene dada por $F_x^{-1}(u) = 2 + 2u$.

Para generar la muestra en cuestión, necesitamos una secuencia de 3 números aleatorios, que podemos obtener usando las técnicas del capítulo 2. Imaginemos que dichos números son 0.3, 0.1 y 0.7. Entonces, la muestra obtenida sería:

$$\begin{aligned} x_1 &= 2 + 2 \times 0.3 = 2.6 \\ x_2 &= 2 + 2 \times 0.1 = 2.2 \\ x_3 &= 2 + 2 \times 0.7 = 3.4 \end{aligned}$$

Generalizando lo anterior a la generación de valores de una variable X con distribución $\mathcal{U}(a, b)$, tenemos que la función de distribución es

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases}$$

Dado $0 < u < 1$, la inversa de la función de distribución viene dada por $F_x^{-1}(u) = a + (b - a)u$, que es la fórmula que utilizaremos en adelante para generar valores de una variable $\mathcal{U}(a, b)$.

Ejemplo 4.2 *Veamos cómo obtener una muestra de una variable exponencial de media β mediante el método de inversión. La densidad de la exponencial de media β es*

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{1}{\beta} e^{-\frac{1}{\beta}x} & \text{si } x > 0 \end{cases}$$

y la función de distribución,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\frac{1}{\beta}x} & \text{si } x > 0 \end{cases}$$

Ahora calculamos la inversa de la función de distribución:

$$\begin{aligned} u = F^{-1}(x) = 1 - e^{-\frac{1}{\beta}x} &\Rightarrow \log e^{-\frac{1}{\beta}x} = \log(1 - u) \\ &\Rightarrow -\frac{1}{\beta}x = \log(1 - u) \\ &\Rightarrow x = -\beta \log(1 - u) . \end{aligned}$$

Ahora bien, si $U \rightsquigarrow \mathcal{U}[0, 1)$ entonces $1 - U \rightsquigarrow \mathcal{U}(0, 1]$ y viceversa, con lo que la inversa quedará como

$$x = -\beta \log u .$$

Como resumen, podemos enunciar un algoritmo para generar una muestra de tamaño n mediante el método de inversión como sigue:

ALGORITMO INVERSIÓN.

1. Desde $i := 1$ hasta n
 - (a) Generar un número aleatorio U .

$$(b) X_i = F_X^{-1}(U).$$

2. Devolver X_1, \dots, X_n como la muestra generada.

La *ventaja* del método de inversión radica en que sólo necesitamos un número aleatorio para obtener un valor de la variable, mientras que su *inconveniente* es que la inversa puede ser difícil de calcular.

4.1.1 Método de inversión para variables discretas

Trataremos en esta sección el caso especial de una variable discreta X que toma un número finito de valores x_1, \dots, x_k . En este caso, la función de distribución de la variable X es escalonada y no siempre estaremos en condiciones de obtener una expresión analítica de la inversa de dicha función de distribución.

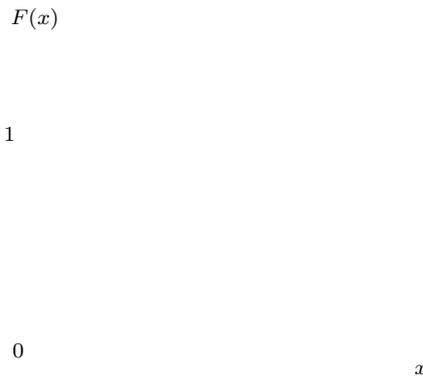


Figura 4.2: Función de distribución de una variable discreta.

La particularización del algoritmo de inversión a esta situación es lo que se conoce como *método de búsqueda en tablas*, ya que el problema de encontrar el inverso de un número aleatorio u se corresponde con buscar el salto correspondiente de la función de distribución (ver figura 4.2), lo que es equivalente a buscar en una tabla de valores de probabilidad acumulados.

Veamos cómo generar una muestra de tamaño n según este método.

ALGORITMO DE BÚSQUEDA EN TABLAS.

1. Desde $i := 1$ hasta n
 - (a) Generar un número aleatorio U .
 - (b) $j := 1$.
 - (c) $acumulado := 0$.
 - (d) $encontrado := falso$.
 - (e) Mientras no encontrado
 - $acumulado := acumulado + P\{X = x_j\}$.
 - Si $U \leq acumulado$
 - $X_i := x_j$.
 - $encontrado := verdadero$.
 - $j := j + 1$.
2. Devolver X_1, \dots, X_n como la muestra generada.

4.2 Método de composición

Pensemos ahora en la posibilidad de que nos encontremos con una variable para la cual no podemos aplicar el método de inversión.

El método de composición se basa en expresar la densidad de la variable a simular como una suma ponderada de densidades para las cuales sea sencillo aplicar el método de inversión. El proceso de simulación se dividirá en dos partes; en primer lugar, se elige de acuerdo con los pesos de cada densidad, una de ellas, y en segundo lugar, se aplica el método de inversión utilizando la función de distribución correspondiente a la densidad elegida.

La justificación de este método viene dada por el siguiente resultado.

Teorema 4.2 *Sea $f_{X|Y}(x|y)$ la función de densidad de una variable X condicionada a otra variable Y con distribución $F_Y(y)$ y densidad $f_Y(y)$. Si simulamos un valor y de*

Y y luego simulamos un valor de X usando la densidad obtenida a partir de $f_{X|Y}$ para $Y = y$, entonces la distribución de los valores de X así obtenidos tiene como densidad

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy .$$

Si Y es discreta con función masa de probabilidad $p_Y(y)$, entonces

$$f_X(x) = \sum_y f_{X|Y}(x|y)p(y) .$$

Demostración: Si simulamos un valor y para la variable Y a partir de $f_Y(y)$ y luego un valor x para la variable X a partir de $f_{X|Y}(x|y)$, estamos simulando valores de la v.a. bidimensional (X, Y) con densidad $f_{X|Y}(x|y) \cdot f_Y(y)$ que es igual a la densidad conjunta $f_{XY}(x, y)$. La distribución de los valores de X así obtenidos podemos calcularla por marginalización:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy .$$

Similar razonamiento se puede hacer en el caso de que Y sea discreta. □

Estudiaremos mediante algunos ejemplos el funcionamiento de este método.

Ejemplo 4.3 Supongamos que queremos obtener una muestra de una variable cuya función de densidad es la que aparece representada en la figura 4.3, cuya expresión es

$$f(x) = \begin{cases} \frac{1}{2}x & \text{si } 0 \leq x \leq 1 \\ \frac{1}{2} & \text{si } 1 < x < 2 \\ \frac{1}{2}(3-x) & \text{si } 2 \leq x \leq 3 \\ 0 & \text{en otro caso.} \end{cases}$$

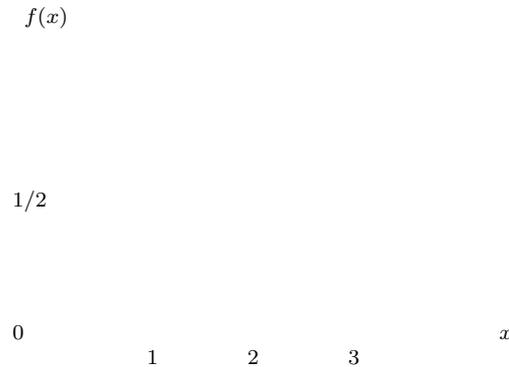


Figura 4.3: Una función de densidad.

Como la densidad aparece definida en tres partes, parece lógico descomponerla como la suma ponderada de tres densidades,

$$f(x) = w_1 f_1(x) + w_2 f_2(x) + w_3 f_3(x) ,$$

donde $w_1 + w_2 + w_3 = 1$. En situaciones como ésta, es muy sencillo obtener tal descomposición, pues cada peso es simplemente el área bajo la correspondiente parte de la función de densidad (en este caso, $w_1 = 1/4$, $w_2 = 1/2$ y $w_3 = 1/4$), y la función de densidad en cada trozo se obtiene simplemente multiplicando la densidad original en esa parte por el inverso del peso. Veámoslo en concreto para cada una de las partes.

La densidad $f_1(x)$ se obtendría multiplicando la densidad $f(x)$ en el intervalo $0 \leq x \leq 1$ por $1/w_1$, es decir,

$$f_1(x) = \frac{1}{1/4} \cdot \frac{1}{2} x = 2x, \quad 0 \leq x \leq 1 .$$

Análogamente,

$$f_2(x) = \frac{1}{1/2} \cdot \frac{1}{2} = 1, \quad 1 < x < 2 ,$$

y

$$f_3(x) = \frac{1}{1/4} \cdot \frac{1}{2}(3-x) = 2(3-x), \quad 2 \leq x \leq 3 .$$

A continuación calculamos las funciones de distribución correspondientes a cada una de las densidades anteriores, y obtenemos sus inversas:

$$F_1(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Y su inversa será

$$F_1^{-1}(u) = \sqrt{u} .$$

De igual manera obtenemos

$$F_2(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ x - 1 & \text{si } 1 < x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

$$F_2^{-1}(u) = u + 1 .$$

Y ya por último,

$$F_3(x) = \begin{cases} 0 & \text{si } x < 2 \\ 1 - (3-x)^2 & \text{si } 2 \leq x \leq 3 \\ 1 & \text{si } x > 3 \end{cases}$$

$$F_3^{-1}(u) = 3 - \sqrt{x} .$$

Con esto, podemos aplicar el método de inversión eligiendo en primer lugar una de las funciones de distribución con probabilidad igual a su peso w_i , $i = 1, 2, 3$, y luego aplicando el método de inversión a la función de distribución elegida. En forma de algoritmo, podemos expresarlo como sigue:

1. Desde $i := 1$ hasta n
 - (a) Generar dos números aleatorios U_1 y U_2 .
 - (b) Si $U_1 \leq 1/4$, $X_i = \sqrt{U_2}$.
 - (c) Si $1/4 < U_1 \leq 3/4$, $X_i = U_2 + 1$.
 - (d) Si $U_1 > 3/4$, $X_i = 3 - \sqrt{U_2}$.
2. Devolver X_1, \dots, X_n como la muestra generada.



Figura 4.4: Una función de densidad.

Ejemplo 4.4 Supongamos que queremos ahora obtener una muestra de una variable cuya densidad es la representada en la figura 4.4, es decir,

$$f(x) = a + 2(1 - a)x, \quad 0 \leq x \leq 1, \quad 0 \leq a \leq 1 .$$

En este caso, la densidad no está definida por trozos como en el ejemplo anterior, por lo que una descomposición como la que hicimos antes no simplifica el problema. En

este caso, podemos pensar en descomponer la densidad “en sentido vertical” en lugar de “horizontal”.

Podemos pensar en descomponer la densidad f en dos partes recorriendo el eje de ordenadas. La primera parte llegaría hasta el valor a y la segunda, desde a hasta $2 - a$.

Esta división proporciona como pesos las áreas de ambas regiones, es decir, $w_1 = a$ y $w_2 = 1 - a$, con lo que las densidades correspondientes son

$$f_1(x) = 1, \quad 0 \leq x \leq 1$$

y

$$f_2(x) = 2x, \quad 0 \leq x \leq 1,$$

donde la descomposición es

$$f(x) = w_1 f_1(x) + w_2 f_2(x).$$

Calculemos ahora las correspondientes funciones de distribución y sus inversas:

$$F_1(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

$$F_1^{-1}(u) = u$$

$$F_2(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

$$F_2^{-1}(u) = \sqrt{u}$$

Podemos enunciar un algoritmo como sigue:

1. Desde $i := 1$ hasta n
 - (a) Generar dos números aleatorios U_1 y U_2 .
 - (b) Si $U_1 \leq a$, $X_i = U_2$.
 - (c) Si $U_1 > a$, $X_i = \sqrt{U_2}$.
2. Devolver X_1, \dots, X_n como la muestra generada.

Como conclusión, hemos de hacer notar que este método, a diferencia del de inversión, necesita dos números aleatorios para obtener un solo valor de la variable. Como contrapartida, se simplifica el cálculo de la inversa de la función de distribución.

4.3 Método de aceptación-rechazo

El método de *aceptación-rechazo* se basa en la utilización de una distribución alternativa para simular un valor de la variable, a lo que sigue la comprobación de cierta condición para ver si el valor generado puede considerarse válido.

Este método se aplicará cuando no sea factible la utilización de las técnicas de inversión y composición.

La idea intuitiva consiste en cubrir la densidad de la variable a simular con una curva de expresión analítica sencilla. A continuación se genera un punto al azar bajo la nueva curva. Si el punto cae por debajo de la densidad original, se considerará válido, y el valor simulado de la variable será la abcisa del punto.

La justificación teórica de este procedimiento viene dada por el siguiente teorema.

Teorema 4.3 *Sea X una variable aleatoria con función de densidad $f_X(x)$, $x \in I \subseteq \mathbb{R}$ tal que puede factorizarse como*

$$f_X(x) = Cg(x)h(x) , \quad (4.1)$$

con $C \in \mathbb{R}$, $C \geq 1$, $0 \leq g(x) \leq 1$ y $h(x)$ función de densidad en I . Sea U una variable con distribución $\mathcal{U}(0,1)$ e Y una variable con densidad $h(y)$ en I . Entonces,

$$f_Y(x|U \leq g(Y)) = f_X(x) . \quad (4.2)$$

Demostración: La densidad condicionada es

$$f_Y(x|U \leq g(Y)) = \frac{P\{U \leq g(Y)|Y = x\}h(x)}{P\{U \leq g(Y)\}} .$$

Ahora bien, por un lado,

$$P\{U \leq g(Y)|Y = x\} = P\{U \leq g(x)\} = g(x) ,$$

dado que U es uniforme en $(0,1)$. Por otro lado,

$$\begin{aligned} P\{U \leq g(Y)\} &= \int_I P\{U \leq g(Y)|Y = x\}h(x)dx \\ &= \int_I g(x)h(x)dx \\ &= \int_I \frac{f_X(x)}{C}dx = \frac{1}{C} , \end{aligned}$$

donde en el penúltimo paso hemos despejado $g(x)h(x)$ de la fórmula (4.1).

Por lo tanto, nos queda que

$$f_Y(x|U \leq g(Y)) = \frac{g(x)h(x)}{1/C} = Cg(x)h(x) = f_X(x) .$$

□

Este teorema quiere decir que si generamos valores de acuerdo con la distribución $h(x)$, los valores generados tendrán la misma distribución que X si en cada uno de ellos se cumple la condición

$$U \leq g(Y) . \quad (4.3)$$

Si esta condición no se cumple, entonces el valor generado será descartado. La probabilidad de que un valor generado sea aceptado es, como vimos en la demostración del teorema anterior,

$$P\{U \leq g(Y)\} = \frac{1}{C} .$$

Llamaremos *eficiencia* del método de aceptación-rechazo al valor $1/C$.

Por tanto, la eficiencia del método de aceptación-rechazo viene determinada por la descomposición que hayamos hecho de la función de densidad original. A la hora de realizar dicha descomposición o factorización, seguiremos los dos siguientes criterios, tratando de llegar a un cierto compromiso entre ambos:

1. Debe ser fácil generar valores a partir de $h(x)$.
2. La eficiencia debe ser lo más alta posible, es decir, C debe estar tan próximo a 1 como se pueda.

En forma de algoritmo, podemos resumir este método como sigue.

ALGORITMO ACEPTACIÓN-RECHAZO.

1. Desde $i := 1$ hasta n
 - (a) válido := falso.
 - (b) Mientras no válido
 - Generar un número aleatorio U .
 - Generar un valor para la variable Y con densidad $h(\cdot)$.

- Si $U \leq g(Y)$
válido := verdadero.

(c) $X_i := Y$.

2. Devolver X_1, \dots, X_n como la muestra generada.

Ejemplo 4.5 Supongamos que queremos obtener una muestra de la variable X que toma valores en el intervalo $[a, b]$ con densidad $f_X(x)$ como la representada en la figura 4.5. En este caso, al tomar la variable valores en un intervalo finito, podemos obtener, de manera sencilla, una descomposición de la densidad $f_X(x)$. En primer lugar, necesitamos definir una función que cubra a dicha densidad (ver figura 4.5). Esa función puede ser

$$t(x) = \max_{a \leq x \leq b} \{f_X(x)\} \quad a \leq x \leq b .$$

A continuación, tomamos una densidad proporcional a $t(x)$ que será la que usemos para simular. Dicha densidad es precisamente la uniforme en el intervalo $[a, b]$ (ver figura 4.5):

$$h(x) = \frac{1}{b-a} \quad a \leq x \leq b .$$

Ahora necesitamos una constante C mayor o igual que 1, que podemos obtener como

$$C = \frac{t(x)}{h(x)} = \frac{t(x)}{1/(b-a)} = t(x) \cdot (b-a) .$$

Por último, obtenemos $g(x)$ despejando de la expresión (4.1) y sustituyendo C y $h(x)$ por su valor:

$$g(x) = \frac{f_X(x)}{Ch(x)} = \frac{f_X(x)}{t(x)(b-a)(1/(b-a))} = \frac{f_X(x)}{t(x)} .$$

Para esta descomposición, el algoritmo de aceptación-rechazo quedaría como sigue:

1. Desde $i := 1$ hasta n
 - (a) válido := falso.
 - (b) Mientras no válido
 - Generar dos números aleatorios U_1 y U_2 .
 - $Y := a + U_2(b - a)$.
 - Si $U_1 \leq \frac{f_x(Y)}{t(Y)}$
válido := verdadero.
 - (c) $X_i := Y$.
2. Devolver X_1, \dots, X_n como la muestra generada.

 $f_X(x)$

$$t(x) = \max\{f_X(x)\}$$

$$h(x) = \frac{1}{b-a}$$

a b x

Figura 4.5: Método de aceptación-rechazo.

Veamos algunos ejemplos de aplicación de este método.

Ejemplo 4.6 *Supongamos que queremos generar una muestra de tamaño 2 para la variable X con densidad*

$$f_X(x) = 3x^2 \quad 0 \leq x \leq 1 \text{ ,}$$

utilizando un generador congruencial lineal módulo 16 y de período máximo.

En primer lugar, hemos de elegir el generador; por ejemplo:

$$x_{n+1} = (5x_n + 3) \pmod{16} \text{ .} \tag{4.4}$$

A continuación, seguimos el mismo proceso que en el ejemplo 4.5 para obtener la descomposición de f_X , es decir:

$$\begin{aligned}t(x) &= 3 \quad 0 \leq x \leq 1 \text{ ,} \\C &= 3 \cdot (b - a) = 3 \cdot (1 - 0) = 3 \text{ ,} \\h(x) &= 1 \quad 0 \leq x \leq 1 \text{ ,} \\g(x) &= \frac{f_X(x)}{t(x)} = \frac{3x^2}{3} = x^2 \quad 0 \leq x \leq 1 \text{ .}\end{aligned}$$

Podemos ver que la eficiencia para esta descomposición es de $1/3$, por lo que el número esperado de pruebas para obtener un resultado válido es 3, siendo la condición de aceptación,

$$U_1 \leq Y^2 \text{ .}$$

Vamos a continuación a obtener la muestra en concreto. Para este fin, necesitamos la secuencia de números aleatorios, que obtendremos a partir del generador (4.4). Si empezamos con semilla 0, los dos primeros números obtenidos serían $U_1 = 0$, $U_2 = 3/16 = 0.1875$, con lo que obtenemos $Y = a + (b - a)U_2 = 0.1875$. Observamos que la condición de aceptación sí se cumple, porque $0 \leq 0.1875^2$, con lo que ya tenemos un elemento de la muestra, $X_1 = 0.1875$.

Ahora calculamos los dos siguientes números que proporciona el generador (4.4), que son $U_1 = 2/16 = 0.125$ y $U_2 = 13/16 = 0.8125$. Con estos números, el valor de Y es 0.8125, con lo que de nuevo se cumple la condición de aceptación y por tanto el segundo elemento de la muestra es $X_2 = 0.8125$.

Veamos algunos ejemplos más donde aplicamos el método de aceptación-rechazo.

Ejemplo 4.7 Supongamos que queremos generar valores para una variable aleatoria X con función de densidad

$$f_X(x) = \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \quad -R \leq x \leq R \text{ .}$$

Como x toma valores en el intervalo finito $[a, b]$, con $a = -R$ y $b = R$, podemos aplicar la misma técnica que en el ejemplo anterior, obteniendo la siguiente descomposición:

$$\begin{aligned} t(x) &= \max_{-R \leq x \leq R} \{f_X(x)\} = \frac{2}{\pi R} \quad -R \leq x \leq R, \\ C &= t(x)(b-a) = \frac{2}{\pi R} 2R = \frac{4}{\pi}, \\ h(x) &= \frac{1}{2R} \quad -R \leq x \leq R, \\ g(x) &= \frac{f_X(x)}{t(x)} = \sqrt{R^2 - x^2} \quad -R \leq x \leq R. \end{aligned}$$

La eficiencia que se obtiene con esta descomposición es

$$\frac{1}{C} = \frac{\pi}{4} \approx 0.785.$$

En resumen, el algoritmo correspondiente quedaría de la siguiente forma:

1. Desde $i := 1$ hasta n .
 - (a) válido := falso.
 - (b) Mientras no válido
 - Generar dos números aleatorios U_1 y U_2 .
 - $Y := -R + U_2(R - (-R)) = -R + U_2 2R = R(2U_2 - 1)$.
 - Si $U_1 \leq \sqrt{R^2 - Y^2}$
válido := verdadero.
 - (c) $X_i := Y$.
2. Devolver X_1, \dots, X_n como la muestra generada.

Ejemplo 4.8 Supongamos que queremos generar valores para una variable aleatoria X con distribución $\beta(4, 3)$, cuya densidad es

$$f_X(x) = \begin{cases} 60x^3(1-x)^2 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Vamos a obtener la descomposición para este caso. En primer lugar, calculamos el máximo de $f_X(x)$ que se alcanza en el punto 0.6 y es igual a 2.0736; es decir,

$$t(x) = 2.0736 \quad 0 \leq x \leq 1 .$$

La función de densidad auxiliar sería

$$h(x) = 1 \quad 0 \leq x \leq 1 ,$$

y además,

$$g(x) = \frac{f_X(x)}{t(x)} = \frac{60x^3(1-x)^2}{2.0736} \quad 0 \leq x \leq 1 .$$

La eficiencia resultante es $1/C = 1/2.0736$, es decir, en promedio necesitaríamos algo más de dos pruebas (4 números aleatorios) para obtener 1 solo valor de la variable X .

4.4 Métodos particulares

Los métodos vistos en las secciones anteriores son de carácter general, debiendo ser adaptados en cada caso a la distribución para la cual deseamos generar valores. En esta sección estudiaremos brevemente la generación de muestras para distribuciones habituales. Remitimos a [7, 10] para un estudio más profundo.

Comenzaremos estudiando algunas variables continuas en primer lugar, dejando las discretas para el final.

4.4.1 Distribuciones Exponencial, Gamma y χ^2

La función de densidad correspondiente a una variable exponencial con media β es

$$f(x) = \frac{1}{\beta} e^{-\frac{1}{\beta}x} \quad x > 0 . \quad (4.5)$$

En el ejemplo 4.2 vimos cómo generar valores de una variable exponencial de media β a partir de números aleatorios, simplemente aplicando la fórmula

$$X := -\beta \log U \quad , \quad (4.6)$$

Por lo tanto, un algoritmo para la distribución exponencial es el siguiente:

ALGORITMO EXPONENCIAL.

1. Desde $i := 1$ hasta n
 - (a) Generar un número aleatorio U_i .
 - (b) $X_i := -\beta \log U_i$.
2. Devolver X_1, \dots, X_n como la muestra generada.

Consideremos ahora el caso de una variable con distribución gamma $\Gamma(k, p)$, cuya densidad es

$$f(x) = \frac{1}{\Gamma(k, p)} x^{k-1} e^{-px} \quad x > 0 \quad , \quad (4.7)$$

donde en $\Gamma(k, p)$ en el denominador se refiere a la función gamma.

Puede observarse que el método de inversión no es directamente aplicable, pues no podemos, en general obtener la expresión analítica de la inversa de la función distribución.

Para resolver este problema, utilizaremos el hecho de que si X_1, \dots, X_k son variables exponenciales independientes con media β , entonces la variable

$$Y = \sum_{i=1}^k X_i$$

sigue una distribución $\Gamma(k, 1/\beta)$. Valiéndonos de este resultado podemos enunciar un sencillo algoritmo:

ALGORITMO GAMMA.

1. Desde $i := 1$ hasta n
 - (a) Generar k números aleatorios U_1, \dots, U_k .
 - (b) $Y_i := -\beta \sum_{j=1}^k \log U_j$.
2. Devolver Y_1, \dots, Y_n como la muestra generada.

En cuanto a la distribución χ_m^2 , donde m son los grados de libertad, recordemos que ésta es igual a una distribución $\Gamma(m/2, 1/2)$. Por lo tanto, en el caso de que m sea par, directamente podemos emplear el algoritmo anterior.

En el caso de que m sea impar, tendremos en cuenta que si Z_i , $i = 1, \dots, m$ son variables independientes con distribución $\mathcal{N}(0, 1)$, entonces

$$X = \sum_{i=1}^m Z_i^2$$

es una variable con distribución χ_m^2 .

En definitiva, si m es impar, para generar un valor de la variable generaremos primero un valor para una $\Gamma((m-1)/2, 1/2)$ (obsérvese que si m es impar entonces $m-1$ es par), luego generamos un valor de una variable $\mathcal{N}(0, 1)$, lo elevamos al cuadrado, y por último sumamos ambos valores, siendo el resultado un valor de la variable χ_m^2 .

4.4.2 Distribución Normal

Dedicaremos especial atención al caso de la distribución normal, dada la gran importancia que tiene dentro de la Estadística.

El objetivo es simular valores de una variable con distribución $\mathcal{N}(\mu, \sigma)$. Para simplificar el problema, nos centraremos en la distribución $\mathcal{N}(0, 1)$, aprovechándonos del hecho de que si $Z \rightsquigarrow \mathcal{N}(0, 1)$, entonces $Y = \sigma Z + \mu$ sigue una distribución $\mathcal{N}(\mu, \sigma)$. Es decir, nos basta generar valores para el caso estándar y luego aplicar esta transformación lineal.

4.4.2.1 Uso del teorema central del límite

Supongamos que disponemos de n variables independientes U_1, \dots, U_n con distribución $\mathcal{U}(0, 1)$. Entonces, de acuerdo con el teorema central del límite, la distribución de

$$Z = \sum_{i=1}^n U_i$$

se aproxima a una distribución normal conforme n crece. Para valores no demasiado altos de n , se obtienen buenas aproximaciones. Por ejemplo, si tomamos $n = 12$, y dado que $E[U_i] = 1/2$ y $\text{Var}(U_i) = 1/12$, entonces la distribución de

$$Z = \frac{\sum_{i=1}^{12} U_i - 12 \cdot 1/2}{\sqrt{12} \sqrt{1/12}} = \sum_{i=1}^{12} U_i - 6 \quad (4.8)$$

es aproximadamente $\mathcal{N}(0, 1)$.

De acuerdo con este fundamento, una forma de obtener un valor de una variable aleatoria con distribución $\mathcal{N}(0, 1)$ es simplemente generar 12 números aleatorios y aplicar la fórmula (4.8).

Hay que ser conscientes de que este esquema no es exacto, por lo que habrá que ser cuidadoso a la hora de su utilización. Por ejemplo, una restricción es que siempre se cumple que

$$-6 \leq Z \leq 6 \text{ ,}$$

lo que puede ser problemático.

En contrapartida, la ventaja de este método radica en su fácil implementación.

En definitiva, podemos especificar un algoritmo para obtener una muestra de tamaño n de una variable aleatoria con distribución $\mathcal{N}(\mu, \sigma)$ como sigue:

ALGORITMO NORMAL-1.

1. Desde $i := 1$ hasta n
 - (a) Generar 12 números aleatorios U_1, \dots, U_n .

$$(b) Z_i := \sum_{j=1}^{12} U_j - 6.$$

$$(c) Y_i := \sigma Z_i + \mu.$$

2. Devolver Y_1, \dots, Y_n como la muestra generada.

4.4.2.2 Método de Box-Muller

A diferencia del anterior, este método es exacto. Se basa en el hecho de que si U_1 y U_2 son dos variables independientes con distribución $\mathcal{U}(0, 1)$, entonces las variables Z_1 y Z_2 definidas como

$$\begin{aligned} Z_1 &= \sqrt{-2 \log U_1} \cos(2\pi U_2) \\ Z_2 &= \sqrt{-2 \log U_1} \operatorname{sen}(2\pi U_2) \end{aligned}$$

son independientes con distribución $\mathcal{N}(0, 1)$. La justificación de este método puede verse, por ejemplo, en [7].

Por tanto, el algoritmo de generación consistiría simplemente en obtener una secuencia de números aleatorios y aplicar las transformaciones anteriores.

Veamos el algoritmo para generar una muestra de tamaño $2n$ de una v.a. con distribución $\mathcal{N}(\mu, \sigma)$:

ALGORITMO NORMAL-2.

1. Desde $i := 1$ hasta n

(a) Generar 2 números aleatorios U_1, U_2 .

(b) $Z_i := \sqrt{-2 \log U_1} \cos(2\pi U_2)$.

(c) $Z_{2i} := \sqrt{-2 \log U_1} \operatorname{sen}(2\pi U_2)$.

(d) $Y_i := \sigma Z_i + \mu$.

(e) $Y_{2i} := \sigma Z_{2i} + \mu$.

2. Devolver Y_1, \dots, Y_{2n} como la muestra generada.

4.4.2.3 Método de aceptación-rechazo para la distribución Normal

Terminaremos el estudio de la distribución normal considerando una descomposición de la densidad de la normal que nos permita aplicar el método de aceptación-rechazo.

Recordemos que la densidad de una variable aleatoria X con distribución normal estándar ($\mathcal{N}(0, 1)$) es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad -\infty < x < \infty . \quad (4.9)$$

Consideremos ahora una variable no negativa cuya densidad sea proporcional a la anterior, es decir,

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2} \quad x > 0 . \quad (4.10)$$

Entonces, puede verse que el problema de generar valores para una $\mathcal{N}(0, 1)$ es equivalente a generar valores a partir de la densidad (4.10) y luego asignarle un signo positivo o negativo al azar. Por tanto, nos centraremos en la densidad (4.10). Recordemos que para aplicar el método de aceptación rechazo tenemos que buscar una descomposición de la función de densidad de la forma

$$f(x) = Ch(x)g(x) .$$

Puede comprobarse que esta descomposición se alcanza para

$$h(x) = e^{-x} \quad (4.11)$$

$$C = \sqrt{\frac{2e}{\pi}} \quad (4.12)$$

$$g(x) = e^{-(x-1)^2/2} \quad (4.13)$$

La eficiencia es

$$\frac{1}{C} = \sqrt{\frac{\pi}{2e}} \approx 0.76 ,$$

y la condición de aceptación, $U \leq g(Y)$, es

$$\begin{aligned} U &\leq e^{-(Y-1)^2/2} \iff \\ -\log U &\geq \frac{(Y-1)^2}{2} , \end{aligned}$$

donde Y sigue una distribución exponencial con media 1. Como $-\log U$ también es exponencial con media 1 si U es un número aleatorio, podemos escribir la condición de aceptación como

$$V_2 \geq \frac{(V_1 - 1)^2}{2} ,$$

donde tanto $V_1 = Y$ como $V_2 = -\log U$ son exponenciales con media 1.

Estamos ya en condiciones de formular el algoritmo detallado.

ALGORITMO NORMAL-3.

1. Desde $i := 1$ hasta n
 - (a) válido := falso.
 - (b) Mientras no válido
 - Generar V_1 y V_2 a partir de la exponencial con media 1.
 - Si $V_2 \geq \frac{(V_1-1)^2}{2}$
válido := verdadero.
 - (c) Generar un número aleatorio U .
 - (d) Si $U \geq 0.5$,
 $Z_i := -V_1$.
En otro caso
 $Z_i := V_1$.
2. Devolver Z_1, \dots, Z_n como la muestra generada.

4.4.3 Distribución Binomial

Consideraremos cómo generar valores de una variable X discreta con distribución binomial $B(n, p)$. Recordemos que la función masa de probabilidad asociada a esta variable es

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n . \quad (4.14)$$

Si n es pequeña, basta con aplicar el algoritmo de inversión descrito en la sección 4.1.1.

En el caso de que n sea grande, el algoritmo de inversión para variables discretas pierde en eficiencia, por lo que podemos pensar en aplicar alguna aproximación que mejore la eficiencia. Concretamente, podemos utilizar el teorema central del límite para aproximar la binomial por una normal, siendo esta aproximación mejor conforme crece el valor de n . Específicamente, la distribución de

$$Z = \frac{X - np + 0.5}{\sqrt{np(1-p)}} \quad (4.15)$$

tiende a ser $\mathcal{N}(0, 1)$ conforme n crece (obsérvese que hemos aplicado la corrección por continuidad).

De manera que para obtener un valor de la variable binomial X , generaremos un valor para Z a partir de la $\mathcal{N}(0, 1)$ y luego despejamos de la ecuación (4.15), teniendo cuidado de redondear siempre a un entero positivo:

$$X = \max \left\{ 0, \left[-0.5 + np + Z \sqrt{np(1-p)} \right] \right\} .$$

4.4.4 Distribución de Poisson

Se dice que una variable aleatoria discreta X sigue una distribución de Poisson si su función masa de probabilidad es

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots \quad (4.16)$$

Para generar valores de una variable de Poisson puede usarse el mismo criterio que para el caso binomial: aplicar el método de inversión para valores pequeños, en estecaso, de λ , y aproximar por una normal para λ grande. De forma más precisa, conforme λ crece, la distribución de

$$Z = \frac{X - \lambda + 0.5}{\sqrt{\lambda}} \quad (4.17)$$

tiende a una $\mathcal{N}(0, 1)$ (obsérvese que hemos aplicado la corrección por continuidad).

Para generar un valor para X , simulamos primero un valor de Z a partir de la $\mathcal{N}(0, 1)$ y luego despejamos de (4.17), redondeando a un entero positivo:

$$X = \max \left\{ 0, \left[\lambda + Z\sqrt{\lambda} - 0.5 \right] \right\} .$$

Capítulo 5

Integración Monte Carlo

En este capítulo estudiaremos los métodos de Monte Carlo mediante la resolución de un problema de carácter determinista, como es estimar el valor de una integral definida. Los métodos que veremos en este capítulo son de utilidad práctica, dado que muchos problemas pueden plantearse como el cálculo de una esperanza (tiempos medios de espera en cola, beneficios, gastos, etc.).

El objetivo es estimar el valor de la integral

$$I = \int_a^b f(x)dx \quad (5.1)$$

Comenzaremos con dos métodos básicos, el de ensayo-error (sección 5.1) y el de la media muestral (5.2). A continuación veremos cómo evaluar la eficiencia del método de Monte Carlo (sección 5.3) y en base a esta evaluación propondremos métodos más sofisticados en la sección 5.4, como el muestreo por importancia, estratificado, sistemático y el uso de variables antitéticas.

5.1 Método de ensayo-error

Se basa en la interpretación geométrica de la integral como un área. Supondremos que $f(x)$ está acotada y es no negativa:

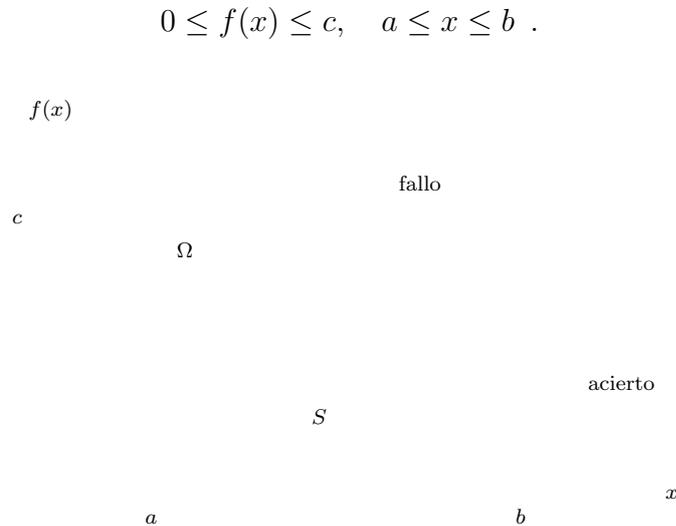


Figura 5.1: Método de ensayo-error.

Sea Ω la región delimitada en la figura 5.1:

$$\Omega = \{(x, y) \mid a \leq x \leq b, \quad 0 \leq y \leq c\} .$$

Denotemos por S la región bajo la curva $f(x)$:

$$S = \{(x, y) \mid y \leq f(x)\} .$$

Obsérvese que el área delimitada por la región S coincide con el valor I .

El método de ensayo-error se basa en generar aleatoriamente puntos dentro de la región Ω y estimar la integral I a partir de la proporción de puntos que caen bajo la curva (es decir, en la región S) de entre el total de puntos generados.

Generar puntos al azar dentro de Ω es lo mismo que obtener una muestra de una variable aleatoria bidimensional (X, Y) con función de densidad

$$f_{XY}(x, y) = \begin{cases} \frac{1}{c(b-a)} & \text{si } (x, y) \in \Omega , \\ 0 & \text{en otro caso.} \end{cases}$$

La probabilidad de que la variable (X, Y) caiga bajo la curva $f(x)$ es

$$p = P\{(X, Y) \in S\} = \frac{\text{área } S}{\text{área } \Omega} = \frac{\int_a^b f(x)dx}{c(b-a)} = \frac{I}{c(b-a)} ,$$

de donde obtenemos que

$$I = c(b-a)p . \quad (5.2)$$

Supongamos que generamos una muestra de tamaño n , (X_i, Y_i) , $i = 1, \dots, n$ de variables con la misma distribución que (X, Y) (uniforme en Ω). Vamos a denominar *acierto* al suceso *un punto muestral cae bajo la curva de $f(x)$* ($f(X_i) \geq Y_i$, $i = 1, \dots, n$), y *fallo* al suceso contrario ($f(X_i) < Y_i$, $i = 1, \dots, n$). Entonces podemos identificar cada punto de la muestra como una prueba de Bernoulli con probabilidad de éxito p .

Si llamamos N_A a la variable *número de aciertos en las n pruebas*, podemos construir un estimador del parámetro p en la fórmula (5.2) como

$$\hat{p} = \frac{N_A}{n} , \quad (5.3)$$

y a partir de él, un estimador de I :

$$\theta_1 = c(b-a) \frac{N_A}{n} . \quad (5.4)$$

En resumen, el procedimiento para estimar I consiste en tomar una muestra de tamaño n a partir de la distribución uniforme en Ω (podemos usar el método de inversión), contar el número de aciertos y evaluar θ_1 .

Ejemplo 5.1 *Vamos a estimar por el método de ensayo-error el valor de la integral*

$$I = \int_0^1 x^3 dx . \quad (5.5)$$

Para ello, realizaremos 4 pruebas ($n = 4$), es decir, necesitaremos 8 números aleatorios, que podemos obtener a partir de un generador. Imaginemos que los números

obtenidos son los siguientes:

$$0.3, 0.5, 0.8, 0.2, 0.7, 0.1, 0.5, 0.6 \ .$$

En primer lugar, debemos determinar la región que vamos a considerar. La más natural viene determinada por $a = 0$, $b = 1$ y $c = 1$, región que encierra la gráfica de la función a integrar. Ahora cogemos los números de dos en dos y calculamos la abscisa (X_i , multiplicando el primer número por $b - a$) y la ordenada (Y_i , multiplicando el segundo número por c) de cada punto, y comprobamos si el punto en cuestión cae por encima o por debajo de la curva.

Para la primera pareja de números, 0.3 y 0.5 , obtenemos $X_i = (1 - 0) \times 0.3 = 0.3$ e $Y_i = 1 \times 0.5 = 0.5$. Para comprobar si se ha producido un acierto o un error, se calcula el valor de la función en X_i y se comprueba si éste está por encima o por debajo de Y_i . En este caso, el valor de la función será $0.3^3 = 0.027$, que es menor que 0.5 , con lo que la prueba ha resultado en un acierto. Con las otras tres parejas obtenemos:

$$\begin{aligned} 0.8^3 = 0.512 > 0.2 &\Rightarrow \text{Fallo.} \\ 0.7^3 = 0.343 > 0.1 &\Rightarrow \text{Fallo.} \\ 0.5^3 = 0.125 < 0.6 &\Rightarrow \text{Acierto.} \end{aligned}$$

Por lo tanto, $N_A = 2$, con lo que $\hat{p} = 2/4 = 0.5$, y el valor estimado será

$$\theta_1 = (1 - 0) \times 1 \times 0.5 = 0.5 \ .$$

Pasemos a estudiar ahora algunas propiedades del estimador θ_1 .

Teorema 5.1 *El estadístico θ_1 definido en la fórmula (5.4) es un estimador insesgado de I con varianza*

$$\text{Var}(\theta_1) = \frac{I}{n}(c(b - a) - I) = \frac{(c(b - a))^2}{n}p(1 - p) \ .$$

Demostración: La variable N_A sigue una distribución Binomial con parámetros n y p . Por tanto,

$$\begin{aligned}
E[\theta_1] &= E\left[c(b-a)\frac{N_A}{n}\right] \\
&= \frac{c(b-a)}{n}E[N_A] \\
&= \frac{c(b-a)}{n}np \\
&= c(b-a)p,
\end{aligned}$$

con lo que $E[\theta_1] = I$, luego θ_1 es un estimador insesgado de I .

Calculemos ahora su varianza.

$$\begin{aligned}
\text{Var}(\theta_1) &= \text{Var}\left(c(b-a)\frac{N_A}{n}\right) \\
&= \frac{(c(b-a))^2}{n^2}\text{Var}(N_A) \\
&= \frac{(c(b-a))^2}{n^2}np(1-p) \\
&= \frac{(c(b-a))^2}{n}p(1-p).
\end{aligned}$$

Si despejamos p de la fórmula (5.2) y sustituimos, obtenemos

$$\begin{aligned}
\text{Var}(\theta_1) &= \frac{(c(b-a))^2}{n} \frac{I}{c(b-a)} \left(1 - \frac{I}{c(b-a)}\right) \\
&= \frac{(c(b-a))^2}{n} \frac{I}{c(b-a)} \left(\frac{c(b-a) - I}{c(b-a)}\right) \\
&= \frac{I}{n}(c(b-a) - I).
\end{aligned}$$

□

Consideraciones sobre el tamaño muestral.

Podemos tratar de establecer, a priori, cuál ha de ser el tamaño muestral para que no sobrepasemos un determinado error ϵ con probabilidad α , es decir, dados $\epsilon, \alpha > 0$, hemos de determinar el valor de n tal que

$$P(|\theta_1 - I| < \epsilon) \geq \alpha .$$

Por la desigualdad de Chebyshev,

$$P(|\theta_1 - I| < \epsilon) \geq 1 - \frac{\text{Var}(\theta_1)}{\epsilon^2} .$$

Para acotar n podemos tomar

$$\begin{aligned} \alpha &\leq 1 - \frac{\text{Var}(\theta_1)}{\epsilon^2} = 1 - \frac{(c(b-a))^2 p(1-p)}{n\epsilon^2} \Leftrightarrow \\ 1 - \alpha &\geq \frac{(c(b-a))^2 p(1-p)}{n\epsilon^2} \Leftrightarrow \\ n &\geq \frac{(c(b-a))^2 p(1-p)}{(1-\alpha)\epsilon^2} . \end{aligned}$$

Como $p(1-p) \leq 1/4$, entonces, en el peor caso obtenemos que

$$n \geq \frac{(c(b-a))^2}{4(1-\alpha)\epsilon^2} .$$

Intervalo de confianza para I .

Podemos construir un intervalo de confianza a nivel $1 - \alpha$ para I aprovechando que para una muestra suficientemente grande, la distribución de

$$\frac{\theta_1 - I}{\sigma_{\theta_1}}$$

es aproximadamente $\mathcal{N}(0, 1)$, donde σ_{θ_1} denota la desviación típica de θ_1 . Veamos:

$$\begin{aligned} z_{\alpha/2} &\leq \frac{\theta_1 - I}{c(b-a)\sqrt{p(1-p)}}\sqrt{n} \leq z_{1-\alpha/2} \\ z_{\alpha/2} \frac{c(b-a)\sqrt{p(1-p)}}{\sqrt{n}} &\leq \theta_1 - I \leq z_{1-\alpha/2} \frac{c(b-a)\sqrt{p(1-p)}}{\sqrt{n}} \end{aligned}$$

$$\theta_1 - z_{1-\alpha/2} \frac{c(b-a)\sqrt{p(1-p)}}{\sqrt{n}} \leq I \leq \theta_1 - z_{\alpha/2} \frac{c(b-a)\sqrt{p(1-p)}}{\sqrt{n}}$$

Como generalmente estamos en condiciones de tomar muestras muy grandes, dado que realizamos los experimentos en un ordenador, podemos sustituir, en la expresión anterior, p por \hat{p} , obteniendo el intervalo aproximado de nivel $1 - \alpha$ para I :

$$\left[\theta_1 - z_{1-\alpha/2} \frac{c(b-a)\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \theta_1 - z_{\alpha/2} \frac{c(b-a)\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \right] \quad (5.6)$$

Ejemplo 5.2 Vamos a calcular un intervalo de confianza al 95% ($1 - \alpha = 0.95$) para el valor de la integral

$$I = \int_0^1 x^3 dx ,$$

usando los datos del ejemplo 5.1. Lo único que nos falta para completar la fórmula (5.6) es $z_{1-\alpha/2}$ y $z_{\alpha/2}$, es decir, $z_{0.975}$ y $z_{0.025}$. Estos valores podemos obtenerlos en las tablas de la normal, y son $z_{0.975} = 1.96$ y $z_{0.025} = -1.96$.

Por tanto, el intervalo que obtenemos sustituyendo en la fórmula (5.6) es

$$[0.01, 0.99] .$$

Obsérvese que el intervalo es muy amplio, debido al pequeño tamaño muestral que hemos utilizado.

A continuación, presentamos el algoritmo detallado que implementa el método de ensayo-error.

ALGORITMO ENSAYO-ERROR

1. $N_A := 0$.
2. Desde $i = 1$ hasta n ,
 - Generar dos números aleatorios U_i y U'_i .

- $X_i := a + U_i(b - a)$.
- Si $f(X_i) > c \cdot U'_i$, hacer $N_A := N_A + 1$.

3. $\hat{p} := N_A/n$.

4. Estimar el valor I como

$$\theta_1 = c(b - a)\hat{p} .$$

5. Dar como varianza de la estimación

$$\hat{\sigma}^2 = (c(b - a))^2 \frac{\hat{p}(1 - \hat{p})}{n} .$$

6. Dar

$$[\theta_1 - z_{1-\alpha/2}\hat{\sigma}, \theta_1 - z_{\alpha/2}\hat{\sigma}]$$

como intervalo de confianza aproximado de nivel $1 - \alpha$ para I .

Obsérvese que no es necesario disponer de manera explícita de la función $f(x)$, sino que basta con poder evaluar si el punto generado cae por debajo de la curva de la función.

5.2 Método de la media muestral

En este caso, la única restricción que imponemos a la función f es:

$$\int_a^b f^2(x)dx < \infty .$$

El método de la media muestral se basa en representar la integral I como el valor esperado de alguna variable aleatoria. Para ello, hacemos una transformación sobre la fórmula (5.1):

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{f^*(x)} f^*(x)dx \tag{5.7}$$

donde $f^*(x)$ es una función de densidad verificando que $f^*(x) > 0$ siempre que $f(x) \neq 0$. Entonces,

$$I = E \left[\frac{f(X)}{f^*(X)} \right] , \quad (5.8)$$

donde X es una variable aleatoria con densidad f^* .

Supongamos por simplicidad que

$$f^*(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, si X es una variable aleatoria con distribución $\mathcal{U}(a, b)$, se cumple que

$$I = (b - a)E[f(X)] . \quad (5.9)$$

Dado que la esperanza puede estimarse mediante la media muestral, para estimar el valor de I basta con generar una muestra de la distribución $\mathcal{U}(a, b)$, evaluar la función f en los puntos de la muestra obtenida y calcular su media. Veamos los detalles.

Sea X_1, X_2, \dots, X_n una muestra de variables con distribución $\mathcal{U}(a, b)$. Podemos definir un estimador del valor I como:

$$\theta_2 = (b - a) \frac{1}{n} \sum_{i=1}^n f(X_i) . \quad (5.10)$$

Ejemplo 5.3 *Vamos a estimar mediante la técnica de la media muestral el valor de la integral (5.5) del ejemplo 5.1, utilizando una muestra de tamaño $n = 4$. Para obtener la muestra necesitamos 4 números aleatorios; tomaremos, por ejemplo, los números*

$$0.3, 0.5, 0.8, 0.2 .$$

Simplemente aplicando la fórmula (5.10) obtenemos la estimación deseada:

$$\theta_2 = (1 - 0) \frac{1}{4} (0.3^3 + 0.5^3 + 0.8^3 + 0.2^3) = 0.168 .$$

Estudiemos a continuación algunas propiedades de θ_2 .

Teorema 5.2 *El estadístico θ_2 definido en la fórmula (5.10) es un estimador insesgado de I con varianza*

$$\text{Var}(\theta_2) = \frac{(b-a)^2}{n} \text{Var}(f(X)) = \frac{1}{n} \left((b-a) \int_a^b f^2(x) dx - I^2 \right) , \quad (5.11)$$

donde $X \rightsquigarrow \mathcal{U}(a, b)$.

Demostración: Veamos en primer lugar que θ_2 es insesgado.

$$\begin{aligned} E[\theta_2] &= E \left[(b-a) \frac{1}{n} \sum_{i=1}^n f(X_i) \right] \\ &= (b-a) E \left[\frac{1}{n} \sum_{i=1}^n f(X_i) \right] \\ &= (b-a) E[f(X)] \\ &= (b-a) \frac{I}{b-a} = I . \end{aligned}$$

A continuación calcularemos su varianza.

$$\begin{aligned} \text{Var}(\theta_2) &= \text{Var} \left((b-a) \frac{1}{n} \sum_{i=1}^n f(X_i) \right) = (b-a)^2 \text{Var} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \right) \\ &= \frac{(b-a)^2}{n} \text{Var}(f(X)) , \end{aligned}$$

donde $X \rightsquigarrow \mathcal{U}(a, b)$. Sustituyendo $\text{Var}(f(X))$ por su valor y empleando la fórmula (5.9),

$$\begin{aligned} \text{Var}(\theta_2) &= \frac{(b-a)^2}{n} (E[f^2(X)] - (E[f(X)])^2) \\ &= \frac{(b-a)^2}{n} \left(\int_a^b f^2(x) \frac{1}{b-a} dx - \frac{I^2}{(b-a)^2} \right) \\ &= \frac{1}{n} \left((b-a) \int_a^b f^2(x) dx - I^2 \right) . \end{aligned}$$

□

Consideraciones sobre el tamaño muestral.

Dados $\epsilon, \alpha > 0$, determinaremos el valor de n tal que

$$P(|\theta_2 - I| < \epsilon) \geq \alpha .$$

Por la desigualdad de Chebyshev,

$$P(|\theta_2 - I| < \epsilon) \geq 1 - \frac{\text{Var}(\theta_2)}{\epsilon^2} ,$$

con lo que garantizamos un error máximo ϵ con probabilidad α si tomamos

$$\begin{aligned} 1 - \frac{\text{Var}(\theta_2)}{\epsilon^2} &\geq \alpha \Leftrightarrow \\ 1 - \frac{(b-a)^2 \text{Var}(f(X))}{n\epsilon^2} &\geq \alpha \Leftrightarrow \\ \frac{(b-a)^2 \text{Var}(f(X))}{n\epsilon^2} &\leq 1 - \alpha \Leftrightarrow \\ \frac{(b-a)^2 \text{Var}(f(X))}{(1-\alpha)\epsilon^2} &\leq n , \end{aligned}$$

donde $X \rightsquigarrow \mathcal{U}(a, b)$.

Para determinar n necesitamos el valor $\text{Var}(f(X))$, que es desconocido. Podemos estimarlo a partir de una *muestra piloto* X_1, \dots, X_n de variables con distribución $\mathcal{U}(a, b)$ como

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (f(X_i) - \bar{f}_n)^2 , \quad (5.12)$$

con

$$\bar{f}_n = \frac{1}{n} \sum_{i=1}^n f(X_i) .$$

Intervalo de confianza para I .

Vamos a denotar por $\sigma_{\theta_2}^2$ la varianza de θ_2 . Por el teorema central del límite, sabemos que la distribución de

$$\frac{\theta_2 - I}{\sqrt{\sigma_{\theta_2}^2}}$$

es aproximadamente $\mathcal{N}(0, 1)$ para un tamaño de muestra n grande.

Para obtener un intervalo de confianza asintótico a nivel $1 - \alpha$ para I tomamos

$$z_{\alpha/2} \leq \frac{\theta_2 - I}{\sqrt{\sigma_{\theta_2}^2}} \leq z_{1-\alpha/2} .$$

Sustituyendo $\sigma_{\theta_2}^2$ por la expresión (5.11), obtenemos

$$z_{\alpha/2} \leq \frac{\theta_2 - I}{\sqrt{\frac{(b-a)^2}{n} \text{Var}(f(X))}} \leq z_{1-\alpha/2} .$$

Como $\text{Var}(f(X))$ es desconocido y n es grande, podemos sustituirlo por s_n^2 , obteniendo:

$$\begin{aligned} z_{\alpha/2} &\leq \frac{\theta_2 - I}{\sqrt{\frac{(b-a)^2}{n} s_n^2}} \leq z_{1-\alpha/2} \Leftrightarrow \\ z_{\alpha/2} &\leq \frac{\theta_2 - I}{(b-a)s_n} \sqrt{n} \leq z_{1-\alpha/2} \Leftrightarrow \\ \theta_2 - z_{1-\alpha/2} \frac{(b-a)s_n}{\sqrt{n}} &\leq I \leq \theta_2 - z_{\alpha/2} \frac{(b-a)s_n}{\sqrt{n}} . \end{aligned}$$

Por tanto, un intervalo de confianza aproximado a nivel $1 - \alpha$ para I viene dado por

$$\left[\theta_2 - z_{1-\alpha/2} \frac{(b-a)s_n}{\sqrt{n}}, \theta_2 - z_{\alpha/2} \frac{(b-a)s_n}{\sqrt{n}} \right] \quad (5.13)$$

Ejemplo 5.4 Calculemos un intervalo de confianza al 95% para la integral del ejemplo 5.3. El único dato que nos falta es la varianza:

$$s_n^2 = \frac{1}{3} \left(\sum_{i=1}^3 f(X_i)^2 - n\bar{f}(X) \right),$$

donde $\bar{f}(X)$ denota la media de los valores de la función para la muestra empleada. Sustituyendo en la expresión anterior, obtenemos

$$s_n^2 = \frac{1}{3}(0.278 - 4 \times 0.168^2) = 0.055 ,$$

con lo que

$$s_n = \sqrt{0.055} = 0.234 .$$

Con esto, obtenemos el intervalo de confianza que buscábamos, sin más que sustituir en la expresión (5.13):

$$\left[0.168 - 1.96 \frac{0.234}{2}, 0.168 + 1.96 \frac{0.234}{2} \right] = [-0.061, 0.397] .$$

Obsérvese que el intervalo es más preciso que el obtenido en el ejemplo 5.2.

A la hora construir el intervalo de confianza anterior, necesitamos calcular el valor s_n^2 . Haremos algunas consideraciones acerca de la implementación del cálculo de dicho valor en un ordenador. La implementación directa de la fórmula, no es la más eficiente, pues requiere explorar todos los valores de la muestra para calcular la media y luego volver a explorarlos para calcular las desviaciones respecto a la media. La fórmula de Köning es más eficiente, pero presenta problemas numéricos (incluso puede llegar a dar valores negativos). Por ello suele utilizarse en la práctica una implementación que se basa en los siguientes resultados:

Proposición 5.1 Para cualesquiera valores reales x_1, \dots, x_n , si denotamos por

$$\bar{x}_k = \frac{1}{k} \sum_{i=1}^k x_i, \quad k = 1, \dots, n ,$$

entonces, para $1 < j \leq n$,

$$\sum_{i=1}^j (x_i - \bar{x}_j)^2 - \sum_{i=1}^{j-1} (x_i - \bar{x}_{j-1})^2 = \frac{j-1}{j} (x_j - \bar{x}_{j-1})^2 .$$

Demostración: Por un lado, aplicando la fórmula de Köning,

$$\begin{aligned} \sum_{i=1}^j (x_i - \bar{x}_j)^2 &= \sum_{i=1}^j x_i^2 - \frac{1}{j} \left(\sum_{i=1}^j x_i \right)^2 \\ &= \sum_{i=1}^j x_i^2 - \frac{1}{j} \left(x_j + \sum_{i=1}^{j-1} x_i \right)^2 \\ &= \sum_{i=1}^j x_i^2 - \frac{1}{j} \left(x_j^2 + \left(\sum_{i=1}^{j-1} x_i \right)^2 + 2x_j \sum_{i=1}^{j-1} x_i \right) \\ &= \sum_{i=1}^j x_i^2 - \frac{1}{j} (x_j^2 + (j-1)^2 \bar{x}_{j-1}^2 + 2x_j(j-1)\bar{x}_{j-1}) . \end{aligned} \quad (5.14)$$

Por otro lado,

$$\sum_{i=1}^{j-1} (x_i - \bar{x}_{j-1})^2 = \sum_{i=1}^{j-1} x_i^2 - \sum_{i=1}^{j-1} \bar{x}_{j-1}^2 + 2\bar{x}_{j-1} \sum_{i=1}^{j-1} x_i . \quad (5.15)$$

Restando (5.15) a (5.14), obtenemos

$$x_j^2 - \frac{x_j^2}{k} - \frac{(j-1)^2}{j} \bar{x}_{j-1}^2 - \frac{j-1}{j} 2x_j \bar{x}_{j-1} - (j-1) \bar{x}_{j-1}^2 + (j-1) 2\bar{x}_{j-1}^2 .$$

Agrupando términos en la expresión anterior, y sacando factor común $(j-1)/j$, vemos que

$$\begin{aligned} \sum_{i=1}^j (x_i - \bar{x}_j)^2 - \sum_{i=1}^{j-1} (x_i - \bar{x}_{j-1})^2 &= \frac{j-1}{j} (x_j^2 + \bar{x}_{j-1}^2 - 2x_j \bar{x}_{j-1}) \\ &= \frac{j-1}{j} (x_j - \bar{x}_{j-1})^2 \end{aligned}$$

□

Teorema 5.3 Dada una muestra X_1, \dots, X_n de variables $\mathcal{U}(a, b)$, para $1 < j \leq n$,

$$(j-1)s_j^2 = (j-2)s_{j-1}^2 + \frac{j-1}{j} (f(X_j) - \bar{f}_{j-1})^2 ,$$

donde

$$\bar{f}_{j-1} = \frac{1}{j-1} \sum_{i=1}^{j-1} f(X_i) .$$

Demostración: Basta con darse cuenta de que

$$(j-1)s_j^2 - (j-2)s_{j-1}^2 = \sum_{i=1}^j (f(X_i) - \bar{f}_j)^2 - \sum_{i=1}^{j-1} (f(X_i) - \bar{f}_{j-1})^2$$

y aplicar la proposición 5.1. □

El teorema anterior proporciona una forma de calcular s_n^2 de forma iterativa, sin necesidad de explorar los valores de la muestra dos veces.

El siguiente algoritmo implementa la estimación de I mediante el método de la media muestral.

ALGORITMO MEDIA MUESTRAL

1. $s := 0, t := 0$.
2. Desde $i = 1$ hasta n
 - Generar un número aleatorio U .
 - $X_i := a + U(b - a)$.
 - Si $i > 1$,

$$t := t + \frac{i-1}{i} \left(f(X_i) - \frac{s}{i-1} \right)^2$$

- $s := s + f(X_i)$.

3. Dar como estimación de I , $\theta_2 = (b - a) \frac{s}{n}$.

$$4. s_n^2 := \frac{t}{n-1}.$$

$$5. \hat{\sigma}_{\theta_2}^2 = \frac{(b-a)^2}{n} s_n^2.$$

6. Dar como intervalo de confianza a nivel $1 - \alpha$ para I ,

$$\left[\theta_2 - z_{1-\alpha/2} \frac{(b-a)s_n}{\sqrt{n}}, \theta_2 - z_{\alpha/2} \frac{(b-a)s_n}{\sqrt{n}} \right].$$

Obsérvese que no es necesario disponer de la expresión explícita de la función $f(x)$ para aplicar este método. Basta con poder evaluarla, por ejemplo, a través de un dispositivo de medida.

5.3 Eficiencia del método de Monte Carlo

Los estimadores construidos en las secciones anteriores son insesgados, por lo que para decidir cuál de los dos es preferible, haremos uso de la varianza. Otro factor que habría que tener en cuenta a la hora de comparar los estimadores, dado que suponemos que estos métodos están orientados a su implementación en un ordenador, es la cantidad de tiempo que se necesita para evaluarlos.

Vamos a estudiar cuál de los dos estimadores vistos, θ_1 y θ_2 es más eficiente, suponiendo que ambos necesitan un tiempo aproximadamente igual para ser evaluados.

Proposición 5.2 θ_2 es más eficiente que θ_1 .

Demostración: Basta con demostrar que $\text{Var}(\theta_2) \leq \text{Var}(\theta_1)$. Veamos:

$$\begin{aligned} \text{Var}(\theta_1) - \text{Var}(\theta_2) &= \left(\frac{I}{n}(c(b-a) - I) \right) - \frac{1}{n} \left((b-a) \int_a^b f^2(x) dx - I^2 \right) \\ &= \frac{I}{n}c(b-a) - \frac{I^2}{n} - \frac{b-a}{n} \int_a^b f^2(x) dx + \frac{I^2}{n} \\ &= \frac{b-a}{n} \left(cI - \int_a^b f^2(x) dx \right). \end{aligned}$$

Como $f(x) \leq c$ para $a \leq x \leq b$, tenemos que

$$cI - \int_a^b f^2(x)dx \geq 0 ,$$

y por tanto, $\text{Var}(\theta_1) - \text{Var}(\theta_2) \geq 0$, con lo que $\text{Var}(\theta_2) \leq \text{Var}(\theta_1)$. \square

5.4 Mejora de la eficiencia

En esta sección estudiaremos algunas posibilidades para mejorar la eficiencia de los estimadores vistos anteriormente. La mejora en la eficiencia vendrá dada en términos de reducción de la varianza.

5.4.1 Muestreo por importancia

Recordemos que el objetivo es estimar el valor

$$I = \int_a^b f(x)dx , \quad (5.16)$$

donde de nuevo supondremos que la integral $\int_a^b f^2(x)dx$ es finita.

La idea consiste en concentrar la distribución de los puntos muestrales en aquellas partes del intervalo (a, b) de mayor importancia en términos de la función f , en lugar de muestrear uniformemente.

Abordaremos el problema en los mismos términos que el método de la media muestral, modificando la fórmula (5.16) de la siguiente manera:

$$I = \int_a^b \frac{f(x)}{f^*(x)} f^*(x)dx = E \left[\frac{f(X)}{f^*(X)} \right] , \quad (5.17)$$

donde X es una v.a. con función de densidad f^* en (a, b) , tal que $f^*(x) > 0$ en todo

punto $x \in (a, b)$ donde $f(x) \neq 0$. A la función f^* se le llama *distribución del muestreo por importancia*.

El proceso de estimación sería análogo al de la media muestral: tomaríamos una muestra X_1, \dots, X_n a partir de f^* y evaluaríamos el estimador

$$\theta_3 = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{f^*(X_i)} . \quad (5.18)$$

Ejemplo 5.5 *Vamos a estimar mediante muestreo por importancia el valor de*

$$I = \int_0^1 x^3 dx ,$$

a partir de una muestra de tamaño $n = 4$ y tomando como distribución de muestreo

$$f^*(x) = 2x \quad 0 < x < 1 .$$

Para tal fin, necesitaremos 4 números aleatorios; por ejemplo,

$$0.3, 0.5, 0.8, 0.2 .$$

Ahora estos números tendríamos que llevarlos al intervalo donde estamos calculando la integral. Como dicho intervalo es el $(0, 1)$, no tenemos que realizar ninguna transformación. La estimación la obtendremos simplemente evaluando la fórmula (5.18):

$$\begin{aligned} \theta_3 &= \frac{1}{4} \left(\frac{0.3^3}{2 \times 0.3} + \frac{0.5^3}{2 \times 0.5} + \frac{0.8^3}{2 \times 0.8} + \frac{0.2^3}{2 \times 0.2} \right) \\ &= \frac{1}{4} (0.045 + 0.125 + 0.32 + 0.02) \\ &= 0.1275 . \end{aligned}$$

A continuación, estudiaremos las propiedades del estimador y veremos también cómo elegir la función de muestreo f^* .

Teorema 5.4 θ_3 es un estimador insesgado de I , con varianza

$$\text{Var}(\theta_3) = \frac{1}{n} \text{Var} \left(\frac{f(X)}{f^*(X)} \right) , \quad (5.19)$$

donde $X \rightsquigarrow f^*$.

Demostración: Veamos en primer lugar que θ_3 es insesgado.

$$\begin{aligned} E[\theta_3] &= E \left[\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{f^*(X_i)} \right] \\ &= \frac{1}{n} n E \left[\frac{f(X)}{f^*(X)} \right] \\ &= E \left[\frac{f(X)}{f^*(X)} \right] \\ &= I , \end{aligned}$$

donde en el último paso hemos usado la fórmula (5.17).

Ahora calculemos la varianza de θ_3 :

$$\text{Var}(\theta_3) = \text{Var} \left(\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{f^*(X_i)} \right) = \frac{1}{n} \text{Var} \left(\frac{f(X)}{f^*(X)} \right) .$$

□

Los siguientes desarrollos nos ayudarán a determinar qué podemos hacer para mejorar la eficiencia.

Proposición 5.3 Sea X una v.a. con densidad $f^*(x)$, $x \in (a, b)$. Sea

$$\xi(X) = \frac{f(X)}{f^*(X)} . \quad (5.20)$$

Entonces,

$$\text{Var}(\xi(X)) = \int_a^b \frac{f^2(x)}{f^*(x)} dx - I^2 .$$

Demostración: En primer lugar, calcularemos la esperanza de ξ :

$$E[\xi(X)] = \int_a^b \frac{f(x)}{f^*(x)} f^*(x) dx = \int_a^b f(x) dx = I .$$

Con esto,

$$\begin{aligned} \text{Var}(\xi(X)) &= E[\xi^2(X)] - (E[\xi(X)])^2 \\ &= \int_a^b \frac{f^2(x)}{(f^*(x))^2} f^*(x) dx - I^2 \\ &= \int_a^b \frac{f^2(x)}{f^*(x)} dx - I^2 \end{aligned}$$

□

A continuación veremos cómo elegir la densidad f^* para minimizar la varianza de $\xi(X)$, que es lo mismo que minimizar la varianza de θ_3 .

Teorema 5.5 *La mínima varianza de $\xi(X)$ es igual a*

$$\text{Var}(\xi_0(X)) = \left(\int_a^b |f(x)| dx \right)^2 - I^2 \quad (5.21)$$

y se alcanza cuando

$$f^*(x) = \frac{|f(x)|}{\int_a^b |f(x)| dx}, \quad a < x < b \quad (5.22)$$

Demostración: En primer lugar veremos que la varianza de ξ cuando f^* viene dada por la expresión (5.22) es efectivamente la que se muestra en la fórmula (5.21). Tenemos que,

$$\begin{aligned}
\text{Var}(\xi_0(X)) &= \int_a^b \frac{f^2(x)}{f^*(x)} dx - I^2 \\
&= \int_a^b \frac{f^2(x)}{|f(x)|} \left(\int_a^b |f(y)| dy \right) dx - I^2 \\
&= \left(\int_a^b |f(y)| dy \right) \int_a^b \frac{f^2(x)}{|f(x)|} dx - I^2 \\
&= \left(\int_a^b |f(x)| dx \right)^2 - I^2 .
\end{aligned}$$

A continuación tenemos que demostrar que $\text{Var}(\xi_0(X)) \leq \text{Var}(\xi(X))$, para lo cual es suficiente probar que:

$$\left(\int_a^b |f(x)| dx \right)^2 \leq \int_a^b \frac{f^2(x)}{f^*(x)} dx .$$

Veamos:

$$\begin{aligned}
\left(\int_a^b |f(x)| dx \right)^2 &= \left(\int_a^b \frac{|f(x)|}{(f^*(x))^{1/2}} (f^*(x))^{1/2} dx \right)^2 \quad [\text{Cauchy-Schwartz}] \\
&\leq \left(\int_a^b \frac{f^2(x)}{f^*(x)} dx \right) \left(\int_a^b f^*(x) dx \right) \\
&= \int_a^b \frac{f^2(x)}{f^*(x)} dx ,
\end{aligned}$$

donde en el último paso hemos tenido en cuenta el hecho de que f^* es densidad en (a, b) y por tanto,

$$\left(\int_a^b f^*(x) dx \right) = 1 .$$

□

Corolario 5.1 Si $f(x) > 0$, $a < x < b$, la distribución del muestreo por importancia óptima es:

$$f^*(x) = \frac{f(x)}{I}, \quad a < x < b , \quad (5.23)$$

y además, en este caso, $\text{Var}(\xi(X)) = 0$.

Demostración: Si $f(x) > 0$, $a < x < b$, entonces la densidad óptima en (a, b) es:

$$f^*(x) = \frac{|f(x)|}{\int_a^b |f(x)| dx} = \frac{f(x)}{\int_a^b |f(x)| dx} = \frac{f(x)}{I} .$$

Por otro lado,

$$\text{Var}(\xi(X)) = \int_a^b \frac{f^2(x)}{f^*(x)} dx - I^2 = I \int_a^b f(x) dx - I^2 = 0 .$$

□

De los resultados anteriores observamos que para usar la distribución del muestreo por importancia óptima, necesitamos conocer el valor

$$\int_a^b |f(x)| dx ,$$

que es prácticamente lo mismo que conocer el valor de la integral I . Por lo tanto, el teorema 5.5 no es de aplicación directa. Sin embargo, nos da una idea sobre la forma en que podemos proceder para reducir la varianza. La densidad óptima es proporcional a la función que se desea integrar, de forma que, en el caso de mínima varianza, el cociente $f(X_i)/f^*(X_i)$ para cada X_i^* de la muestra, también llamado *peso* de X_i , se mantiene constante. Una forma de proceder sería tomar f^* tan próxima a f en forma como sea posible, de cara a obtener unos pesos más uniformes.

Consideraciones sobre el tamaño muestral.

Dados $\epsilon, \alpha > 0$, determinaremos el valor de n tal que

$$P(|\theta_3 - I| < \epsilon) \geq \alpha .$$

Por la desigualdad de Chebyshev,

$$P(|\theta_3 - I| < \epsilon) \geq 1 - \frac{\text{Var}(\theta_3)}{\epsilon^2} ,$$

con lo que garantizamos un error máximo ϵ con probabilidad α si tomamos

$$\begin{aligned}
1 - \frac{\text{Var}(\theta_3)}{\epsilon^2} &\geq \alpha \Leftrightarrow \\
1 - \frac{\text{Var}(\xi(X))}{n\epsilon^2} &\geq \alpha \Leftrightarrow \\
\frac{\text{Var}(\xi(X))}{n\epsilon^2} &\leq 1 - \alpha \Leftrightarrow \\
\frac{\text{Var}(\xi(X))}{(1 - \alpha)\epsilon^2} &\leq n ,
\end{aligned}$$

donde $X \rightsquigarrow f^*$.

Para determinar n necesitamos el valor $\text{Var}(\xi(X))$, que es desconocido. Podemos estimarlo a partir de una *muestra piloto* X_1, \dots, X_n de variables con distribución f^* como

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi(X_i) - \bar{\xi}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{f(X_i)}{f^*(X_i)} - \bar{\xi}_n \right)^2 , \quad (5.24)$$

donde

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi(X_i) = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{f^*(X_i)} .$$

Intervalo de confianza para I .

Vamos a denotar por $\sigma_{\theta_3}^2$ la varianza de θ_3 . Por el teorema central del límite, sabemos que la distribución de

$$\frac{\theta_3 - I}{\sqrt{\sigma_{\theta_3}^2}}$$

es aproximadamente $\mathcal{N}(0, 1)$ para un tamaño de muestra n grande.

Para obtener un intervalo de confianza asintótico a nivel $1 - \alpha$ para I tomamos

$$z_{\alpha/2} \leq \frac{\theta_3 - I}{\sqrt{\sigma_{\theta_3}^2}} \leq z_{1-\alpha/2} .$$

Sustituyendo $\sigma_{\theta_2}^2$ por la expresión (5.19), obtenemos

$$z_{\alpha/2} \leq \frac{\theta_3 - I}{\sqrt{\frac{1}{n} \text{Var}(\xi(X))}} \leq z_{1-\alpha/2} .$$

Como $\text{Var}(\xi(X))$ es desconocido y n es grande, podemos sustituirlo por s_n^2 , obteniendo:

$$\begin{aligned} z_{\alpha/2} &\leq \frac{\theta_3 - I}{\sqrt{\frac{1}{n} s_n^2}} \leq z_{1-\alpha/2} \Leftrightarrow \\ z_{\alpha/2} &\leq \frac{\theta_3 - I}{s_n} \sqrt{n} \leq z_{1-\alpha/2} \Leftrightarrow \\ \theta_3 - z_{1-\alpha/2} \frac{s_n}{\sqrt{n}} &\leq I \leq \theta_3 - z_{\alpha/2} \frac{s_n}{\sqrt{n}} . \end{aligned}$$

Por tanto, un intervalo de confianza aproximado a nivel $1 - \alpha$ para I viene dado por

$$\left[\theta_3 - z_{1-\alpha/2} \frac{s_n}{\sqrt{n}}, \theta_3 - z_{\alpha/2} \frac{s_n}{\sqrt{n}} \right] \quad (5.25)$$

Ejemplo 5.6 Continuaremos el ejemplo 5.5 para obtener un intervalo de confianza al 95% para I . Sólo nos falta calcular s_n para completar la fórmula (5.25).

$$\begin{aligned} s_n^2 &= \frac{1}{3}(0.045^2 + 0.125^2 + 0.32^2 + 0.02^2 - 4 \times 0.1275^2) \\ &= 0.018 , \end{aligned}$$

con lo que

$$s_n = \sqrt{0.018} = 0.13 .$$

Por lo tanto, el intervalo que buscábamos es:

$$\left[0.1275 - 1.96 \frac{0.13}{2}, 0.1275 + 1.96 \frac{0.13}{2} \right] = [0.0001, 0.2549] .$$

El siguiente algoritmo implementa la estimación de I mediante la técnica del muestreo por importancia.

ALGORITMO MUESTREO POR IMPORTANCIA

1. Seleccionar la distribución del muestreo por importancia, f^* .
2. $s := 0, t := 0$.
3. Desde $i = 1$ hasta n
 - Generar un número aleatorio U .
 - Obtener un valor X_i a partir de U aplicando el método de inversión para la distribución del muestreo por importancia.
 - $\xi := \frac{f(X_i)}{f^*(X_i)}$.
 - Si $i > 1$,

$$t := t + \frac{i-1}{i} \left(\xi - \frac{s}{i-1} \right)^2$$

- $s := s + \xi$.
4. Dar como estimación de I , $\theta_3 = \frac{s}{n}$.
 5. $s_n^2 := \frac{t}{n-1}$.
 6. $\hat{\sigma}_{\theta_3}^2 = \frac{s_n^2}{n}$.
 7. Dar como intervalo de confianza a nivel $1 - \alpha$ para I ,

$$\left[\theta_3 - z_{1-\alpha/2} \frac{s_n}{\sqrt{n}}, \theta_3 + z_{\alpha/2} \frac{s_n}{\sqrt{n}} \right] .$$

5.4.2 Muestreo estratificado

En esta ocasión, de nuevo expresamos el valor I como:

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{f^*(x)} f^*(x)dx = \int_a^b g(x)f^*(x)dx , \quad (5.26)$$

donde f^* es una función de densidad en (a, b) en las mismas condiciones que en el muestreo por importancia, y

$$g(x) = \frac{f(x)}{f^*(x)} \quad a < x < b .$$

En el muestreo estratificado, el esfuerzo no se concentra en la selección de la función f^* , sino que la reducción de la varianza viene dada por la división del intervalo (a, b) en una serie de regiones disjuntas en cada una de las cuales se tomará un determinado número de puntos muestrales. Pasamos a analizar en detalle este método.

Consideremos una partición del intervalo $D = (a, b)$: $D_i, i = 1, \dots, m, D = \cup_{i=1}^m D_i, D_j \cap D_k = \emptyset$ si $j \neq k$. La integral de la función f en cada región D_i es

$$I_i = \int_{D_i} g(x)f^*(x)dx , \quad (5.27)$$

que podemos estimar, por ejemplo, mediante la técnica de la media muestral.

Definimos

$$P_i = \int_{D_i} f^*(x)dx, \quad i = 1, \dots, m . \quad (5.28)$$

Se cumple que $\sum_{i=1}^m P_i = 1$ y podemos expresar la integral I como

$$\begin{aligned} I &= \int_D g(x)f^*(x)dx \\ &= \sum_{i=1}^m \int_{D_i} g(x)f^*(x)dx = \sum_{i=1}^m I_i . \end{aligned}$$

Si tomamos

$$g_i(x) = \begin{cases} g(x) & \text{si } x \in D_i \\ 0 & \text{si } x \notin D_i \end{cases} \quad (5.29)$$

podemos construir una función de densidad para cada región D_i simplemente dividiendo f^* entre P_i ,

$$f_i^*(x) = \frac{f^*(x)}{P_i} \quad x \in D_i, \quad i = 1, \dots, m, \quad (5.30)$$

dado que

$$\int_{D_i} \frac{f^*(x)}{P_i} dx = \frac{1}{P_i} \int_{D_i} f^*(x) dx = \frac{P_i}{P_i} = 1 .$$

Teniendo en cuenta lo anterior, podemos escribir I_i , $i = 1, \dots, m$, multiplicando y dividiendo por P_i , como

$$\begin{aligned} I_i &= \int_{D_i} g(x) \frac{f^*(x)}{P_i} P_i dx \\ &= P_i \int_{D_i} g(x) \frac{f^*(x)}{P_i} dx \\ &= P_i \int_{D_i} g_i(x) f_i^*(x) dx \\ &= P_i E [g_i(X_i)] , \end{aligned}$$

siendo X_i una v.a. con densidad $f_i^*(x)$, $x \in D_i$.

La esperanza en la expresión anterior podemos estimarla mediante la media muestral, con lo que un estimador insesgado para cada integral I_i viene dado por

$$\tau_i = \frac{P_i}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}), \quad i = 1, \dots, m, \quad (5.31)$$

donde $\{X_{k_i}\}$, $k_i = 1, \dots, n_i$, es una muestra con densidad $f_i^*(x)$, $x \in D_i$, y n_i es el tamaño de la muestra, es decir, la cantidad de observaciones que se toman en la región D_i .

De esta manera, podemos definir el estimador de muestreo estratificado del parámetro I como:

$$\theta_4 = \sum_{i=1}^m \tau_i = \sum_{i=1}^m \frac{P_i}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}) , \quad (5.32)$$

donde el tamaño total de la muestra es

$$n = \sum_{i=1}^m n_i .$$

Veamos a continuación algunas propiedades del estimador θ_4 .

Teorema 5.6 θ_4 es un estimador insesgado de I con varianza

$$\text{Var}(\theta_4) = \sum_{i=1}^m \frac{P_i^2}{n_i} \text{Var}(g(X_i)) = \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} , \quad (5.33)$$

donde X_i , $i = 1, \dots, m$ son v.a. con densidad f_i^* , y

$$\sigma_i^2 = \text{Var}(g_i(X_i)) = \frac{1}{P_i} \int_{D_i} g_i^2(x) f_i^*(x) dx - \frac{I_i^2}{P_i^2} .$$

Demostración: Veamos en primer lugar que θ_4 es insesgado:

$$E[\theta_4] = E \left[\sum_{i=1}^m \tau_i \right] = \sum_{i=1}^m E[\tau_i] = \sum_{i=1}^m I_i = I ,$$

donde hemos utilizado que τ_i es un estimador insesgado de I_i , $i = 1, \dots, m$.

Calculemos a continuación su varianza.

$$\text{Var}(\theta_4) = \text{Var} \left(\sum_{i=1}^m \frac{P_i}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}) \right) .$$

Dado que las observaciones en cada región son independientes, podemos descomponer la varianza de la suma como la suma de las varianzas:

$$\begin{aligned} \sum_{i=1}^m \text{Var} \left(\frac{P_i}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}) \right) &= \sum_{i=1}^m P_i^2 \text{Var} \left(\frac{1}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}) \right) \\ &= \sum_{i=1}^m \frac{P_i^2}{n_i} \text{Var} (g(X_{k_i})) \\ &= \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} . \end{aligned}$$

Por otro lado,

$$\begin{aligned} \sigma_i^2 &= \text{Var} (g_i(X_i)) \\ &= E [g_i^2(X_i)] - (E [g_i(X_i)])^2 \\ &= \int_{D_i} g_i^2(X_i) f_i^*(x) dx - \frac{I_i^2}{P_i^2} \\ &= \int_{D_i} g_i^2(X_i) \frac{f_i^*(x)}{P_i} dx - \frac{I_i^2}{P_i^2} \\ &= \frac{1}{P_i} \int_{D_i} g_i^2(X_i) f_i^*(x) dx - \frac{I_i^2}{P_i^2} , \end{aligned}$$

donde X_i es una v.a. con densidad f_i^* en D_i , $i = 1, \dots, m$. □

A continuación estudiaremos cómo estratificar de forma óptima.

Teorema 5.7 *Sea $\{D_i\}$, $i = 1, \dots, m$, una partición del intervalo $D = (a, b)$. Entonces,*

$$\min \{ \text{Var}(\theta_4) \} = \frac{1}{n} \left(\sum_{i=1}^m P_i \sigma_i \right)^2 ,$$

con $\sum_{i=1}^m n_i = n$, y ocurre cuando

$$n_i = n \frac{P_i \sigma_i}{\sum_{j=1}^m P_j \sigma_j} . \quad (5.34)$$

Demostración: Tenemos que

$$\text{Var}(\theta_4) = \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} .$$

Sustituyendo n_i por (5.34) obtenemos

$$\begin{aligned} \text{Var}(\theta_4) &= \sum_{i=1}^m \frac{P_i^2}{n P_i \sigma_i} \left(\sum_{j=1}^m P_j \sigma_j \right) \sigma_i^2 \\ &= \frac{1}{n} \left(\sum_{i=1}^m P_i \sigma_i \left(\sum_{j=1}^m P_j \sigma_j \right) \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^m P_i \sigma_i \right) \left(\sum_{j=1}^m P_j \sigma_j \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^m P_i \sigma_i \right)^2 . \end{aligned}$$

Ahora hemos de ver que éste es el mínimo valor de θ_4 :

$$\begin{aligned} \frac{1}{n} \left(\sum_{i=1}^m P_i \sigma_i \right)^2 &= \frac{1}{n} \left(\sum_{i=1}^m \frac{P_i \sigma_i}{\sqrt{n_i}} \sqrt{n_i} \right)^2 \quad [\text{Cauchy-Schwartz}] \\ &\leq \frac{1}{n} \left(\sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} \right) \left(\sum_{i=1}^m n_i \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} \right) n \\ &= \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i} , \end{aligned}$$

que es el valor de la varianza de θ_4 . □

Proposición 5.4 *Sea una estratificación D_i , $i = 1, \dots, m$ con $n_i = nP_i$. Entonces,*

$$\text{Var}(\theta_4) \leq \text{Var}(\theta_3) .$$

Demostración:

$$\begin{aligned} \text{Var}(\theta_4) &= \sum_{i=1}^m \frac{P_i^2 \text{Var}(g_i(X_i))}{n_i} \\ &= \sum_{i=1}^m \frac{P_i^2 \text{Var}(g_i(X_i))}{nP_i} \\ &= \frac{1}{n} \sum_{i=1}^m P_i \text{Var}(g_i(X_i)) . \end{aligned}$$

De acuerdo con la desigualdad de Cauchy-Schwartz,

$$I^2 = \left(\sum_{i=1}^m I_i \right)^2 = \left(\sum_{i=1}^m \frac{I_i}{\sqrt{P_i}} \sqrt{P_i} \right)^2 \leq \left(\sum_{i=1}^m \frac{I_i^2}{P_i} \right) \left(\sum_{i=1}^m P_i \right) = \left(\sum_{i=1}^m \frac{I_i^2}{P_i} \right) .$$

Ahora, multiplicando por P_i y sumando en i en la fórmula de la varianza de $g_i(X_i)$,

$$\text{Var}(g_i(X_i)) = \frac{1}{P_i} \int_{D_i} g_i^2(x) f^*(x) dx - \frac{I_i^2}{P_i} ,$$

obtenemos

$$\begin{aligned} \sum_{i=1}^m P_i \text{Var}(g_i(X_i)) &= \int_D g^2(x) f^*(x) dx - \sum_{i=1}^m \frac{I_i^2}{P_i} \\ &\leq \int_D g^2(x) f^*(x) dx - I^2 \\ &= n \text{Var}(\theta_3) . \end{aligned}$$

Con lo que $n \text{Var}(\theta_4) \leq n \text{Var}(\theta_3)$, o lo que es lo mismo, $\text{Var}(\theta_4) \leq \text{Var}(\theta_3)$. □

Intervalo de confianza para I .

Vamos a denotar por $\sigma_{\theta_4}^2$ la varianza de θ_4 . Por el teorema central del límite, sabemos que la distribución de

$$\frac{\sum_{i=1}^m \tau_i - \sum_{i=1}^m E[\tau_i]}{\sqrt{\sum_{i=1}^m \text{Var}(\tau_i)}} = \frac{\theta_4 - I}{\sqrt{\sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{n_i}}}$$

es aproximadamente $\mathcal{N}(0, 1)$ para m grande.

Los valores σ_i^2 en la expresión anterior son desconocidos, por lo que podemos aproximarlos por

$$s_i^2 = \frac{1}{n_i - 1} \sum_{k_i=1}^{n_i} (g(X_{k_i}) - \bar{g}_i)^2, \quad (5.35)$$

donde

$$\bar{g}_i = \frac{1}{n_i} \sum_{k_i=1}^{n_i} g(X_{k_i}).$$

Para obtener un intervalo de confianza asintótico a nivel $1 - \alpha$ para I tomamos

$$z_{\alpha/2} \leq \frac{\theta_4 - I}{\sqrt{\sum_{i=1}^m \frac{P_i^2 s_i^2}{n_i}}} \leq z_{1-\alpha/2}.$$

Por tanto, un intervalo de confianza aproximado a nivel $1 - \alpha$ para I viene dado por

$$\left[\theta_4 - z_{1-\alpha/2} \sqrt{\sum_{i=1}^m \frac{P_i^2 s_i^2}{n_i}}, \theta_4 - z_{\alpha/2} \sqrt{\sum_{i=1}^m \frac{P_i^2 s_i^2}{n_i}} \right] \quad (5.36)$$

El siguiente algoritmo implementa la estimación de I mediante la técnica del muestreo estratificado.

ALGORITMO MUESTREO ESTRATIFICADO

1. Seleccionar la distribución de muestreo f^* y una partición $\{D_i\}$, $i = 1, \dots, m$ del intervalo (a, b) .
2. $s := 0$, $v := 0$.
3. Desde $i = 1$ hasta m ,
 - (a) $s_i := 0$, $t_i := 0$.
 - (b) Desde $k_i = 1$ hasta n_{k_i} ,
 - Generar un número aleatorio U .
 - Obtener un valor X_{k_i} a partir de U aplicando el método de inversión para la distribución de muestreo.
 - Si $k_i > 1$,

$$t_i := t_i + \frac{k_i - 1}{k_i} \left(g(X_{k_i}) - \frac{s_i}{k_i - 1} \right)^2$$
 - $s_i := s_i + g(X_{k_i})$.
 - (c) $\tau_i := \frac{P_i s_i}{n_i}$.
 - (d) $s := s + \tau_i$.
 - (e) $\hat{\sigma}_i^2 := \frac{t_i}{n - 1}$.
 - (f) $v := v + \frac{P_i^2 \hat{\sigma}_i^2}{n_i}$.

4. Dar como estimación de I , $\theta_4 = s$.
5. $\hat{\sigma}_{\theta_4}^2 = v$.
6. Dar como intervalo de confianza a nivel $1 - \alpha$ para I ,

$$\left[\theta_4 - z_{1-\alpha/2} \sqrt{\hat{\sigma}_{\theta_4}^2}, \theta_4 - z_{\alpha/2} \sqrt{\hat{\sigma}_{\theta_4}^2} \right] .$$

5.4.3 Muestreo sistemático

Se llama *muestreo sistemático* a un caso particular de muestreo estratificado en el que todas las regiones tienen igual probabilidad y se toma idéntico número de observaciones en cada una de ellas; es decir, si n es el tamaño de la muestra y m el número de regiones,

$$P_i = \frac{1}{m} \quad \text{y} \quad n_i = \frac{n}{m} \quad i = 1, \dots, m .$$

Con estas restricciones, podemos enunciar el algoritmo de muestreo sistemático como sigue:

ALGORITMO MUESTREO SISTEMÁTICO

1. Seleccionar la distribución de muestreo f^* .
2. Dividir el rango $[0, 1]$ de la función de distribución de f^* en m intervalos de amplitud $1/m$ cada uno, obteniendo así m regiones D_i , $i = 1, \dots, m$ cada una con probabilidad $P_i = 1/m$.
3. $s := 0$, $v := 0$.
4. Desde $i = 1$ hasta m ,
 - (a) $s_i := 0$, $t_i := 0$.
 - (b) Desde $k_i = 1$ hasta n/m ,
 - Generar un número aleatorio U .
 - Llevar el número U a la región D_i mediante

$$U' = \frac{i - 1 + U}{m} . \tag{5.37}$$

- Obtener un valor X_{k_i} a partir de U' aplicando el método de inversión para la distribución de muestreo.
- Si $k_i > 1$,

$$t_i := t_i + \frac{k_i - 1}{k_i} \left(g(X_{k_i}) - \frac{s_i}{k_i - 1} \right)^2$$

$$\begin{aligned}
 & \bullet s_i := s_i + g(X_{k_i}). \\
 \text{(c)} \quad \tau_i &:= \frac{P_i s_i}{n_i} = \frac{P_i s_i m}{n} . \\
 \text{(d)} \quad s &:= s + \tau_i. \\
 \text{(e)} \quad \hat{\sigma}_i^2 &:= \frac{t_i}{n-1}. \\
 \text{(f)} \quad v &:= v + \frac{P_i^2 \hat{\sigma}_i^2}{n_i} = v + \frac{P_i^2 \hat{\sigma}_i^2 m}{n}.
 \end{aligned}$$

5. Dar como estimación de I , $\theta_4 = s$.

6. $\hat{\sigma}_{\theta_4}^2 = v$.

7. Dar como intervalo de confianza a nivel $1 - \alpha$ para I ,

$$\left[\theta_4 - z_{1-\alpha/2} \sqrt{\hat{\sigma}_{\theta_4}^2}, \theta_4 - z_{\alpha/2} \sqrt{\hat{\sigma}_{\theta_4}^2} \right] .$$

Ejemplo 5.7 Para ilustrar el funcionamiento de este algoritmo, vamos a estimar el valor de

$$I = \int_0^1 x dx ,$$

tomando una muestra de tamaño 4 con dos regiones, mediante el método de muestreo sistemático. Vamos a usar como función de muestreo

$$f^*(x) = 1 \quad 0 < x < 1 .$$

Necesitaremos 4 números aleatorios, que podemos obtener mediante un generador. Supongamos que los números obtenidos son 0.7, 0.2, 0.5 y 0.8.

Ahora transformamos los números anteriores de acuerdo con la fórmula (5.37) de manera que nos queden dos números en cada región. Los números transformados son:

$$\frac{0 + 0.7}{2} = \mathbf{0.35}$$

$$\frac{0 + 0.2}{2} = \mathbf{0.1}$$

para la primera región, y

$$\frac{1 + 0.5}{2} = \mathbf{0.75}$$

$$\frac{1 + 0.8}{2} = \mathbf{0.9}$$

para la segunda región.

Aplicando ahora el método de inversión, la muestra obtenida es $\mathbf{X} = \{0.35, 0.1, 0.75, 0.9\}$, y la estimación de I es

$$\theta_4 = \frac{0.35 + 0.1 + 0.75 + 0.9}{4} = 0.525 .$$

5.4.4 Variables antitéticas

Son muchas las estrategias que pueden emplearse con el propósito de la mejora de la eficiencia. En las secciones anteriores hemos visto quizás las de mayor ámbito de aplicación. Terminaremos presentando otra estrategia, también bastante general, y que en ocasiones puede combinarse con las anteriores. Se basa en el uso de variables *antitéticas*, es decir, variables entre las que existe correlación negativa.

Imaginemos que queremos estimar el valor de cierta integral, I , para lo que disponemos de dos estimadores insesgados Y_1 e Y_2 . Tenemos que

$$\hat{I} = \frac{1}{2}(Y_1 + Y_2)$$

es un estimador insesgado de I con varianza

$$\text{Var} \left(\frac{1}{2}(Y_1 + Y_2) \right) = \frac{1}{4}\text{Var}(Y_1) + \frac{1}{4}\text{Var}(Y_2) + \frac{1}{2}\text{Cov}(Y_1, Y_2) .$$

Si la covarianza es fuertemente negativa, puede reducirse en gran medida la varianza del estimador. Veamos un ejemplo de aplicación de esta técnica.

Ejemplo 5.8 Queremos estimar el valor de

$$I = \int_0^1 f(x) dx .$$

Ahora bien, se cumple que

$$I = \int_0^1 f(x)dx = \frac{1}{2} \int_0^1 (f(x) + f(1-x))dx .$$

Si consideramos $U \rightsquigarrow \mathcal{U}(0,1)$ y definimos

$$Y = \frac{1}{2}(Y_1 + Y_2) = \frac{1}{2}(f(U) + f(1-U)) ,$$

donde $Y_1 = f(U)$ y $Y_2 = f(1-U)$, se cumple que

$$E[Y] = I ,$$

por lo que podemos construir un estimador insesgado del parámetro I , para una secuencia de n números aleatorios U_1, \dots, U_n , de la siguiente forma:

$$\theta_5 = \frac{1}{2n} \sum_{i=1}^n (f(U_i) + f(1-U_i)) .$$

El interés de aplicar esta técnica vendrá determinado por la rentabilidad de evaluar el doble de veces la función f respecto, por ejemplo, al método de la media muestral. Concretamente, será rentable si se cumple que

$$\text{Var}(\theta_5) \leq \frac{1}{2} \text{Var}(\theta_2) .$$

Capítulo 6

Técnicas de remuestreo

Hasta el momento hemos tratado dos aspectos fundamentales de la Estadística Computacional. Por un lado, hemos visto cómo generar una muestra aleatoria de una población cuando se conoce cuál es su distribución (capítulo 4).

A continuación estudiamos la posibilidad de estimar el valor de un parámetro de una distribución, siempre que tengamos un mecanismo para generar una muestra artificial. También vimos cómo problemas de carácter determinista pueden ser resueltos de forma aproximada introduciendo un factor de aleatorización, empleando el llamado método de Monte Carlo.

En este capítulo nos centraremos en un problema que, al igual que los tratados hasta el momento, requieren el uso de ordenadores debido a la cantidad de cálculos que conllevan. Básicamente, el objetivo de las técnicas de remuestreo es el de mejorar la “calidad” de una muestra.

6.1 Remuestreo

Podemos esquematizar el problema a tratar en los siguientes puntos:

- Disponemos de una muestra de una población con distribución desconocida, de la que queremos estimar un parámetro θ .
- Para ello construimos un estimador $\hat{\theta}$ de θ .

- Queremos estimar la bondad de la estimación mediante el **error estándar**:
 - Para obtenerlo de forma exacta necesitamos conocer la distribución de la población.
 - En algunos casos puede estimarse.

6.1.1 Error estándar de la media

Particularizaremos el problema presentado al caso de la estimación de la media de cierta distribución. Sea X una v.a. con $E[X] = \mu$ y $\text{Var}(X) = \sigma^2$. Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra de variables i.i.d. con la misma distribución que X . Se define el estimador media muestral como

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

\bar{x} es un estimador insesgado de $E[X]$ con $E[\bar{x}] = \mu$ y $\text{Var}(\bar{X}) = \sigma^2/n$.

Se define el *error estándar* de un estimador $\hat{\theta}$ de un parámetro θ como

$$\text{se}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})} .$$

En el caso particular de la media muestral,

$$\text{se}(\bar{X}) = \frac{\sigma}{\sqrt{n}} .$$

El problema es que para calcularlo necesitamos conocer la varianza del estimador, que es lo mismo que conocer la varianza de X .

Si ésta es desconocida, podemos pensar en utilizar una estimación de la varianza:

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} ,$$

con lo que podemos formular el *error estándar estimado* como

$$\hat{s}e = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{1}{n^2} \sum_{i=1}^n (X_i - \bar{X})^2} .$$

Hemos de notar que este proceso es válido sólo para el estimador media muestral. Necesitamos un procedimiento genérico que permita estimar el error estándar de cualquier estimador.

Obsérvese que estamos tratando de determinar la bondad de la estimación a partir de la misma muestra con la que hemos estimado. Si la estimación fue mala, debido a que la muestra era sesgada, también tendríamos motivos para sospechar de la estimación del error estándar proporcionada por dicha muestra.

Vamos a presentar brevemente dos procedimientos que tratan de solventar este inconveniente, el *bootstrap* y el *jackknife*.

6.2 Bootstrap

Definición 6.1 Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra de v.a. *i.i.d.* cuya distribución es desconocida. Deseamos estimar un parámetro θ mediante un estimador $\hat{\theta}$. Sea \hat{F} una distribución de probabilidad que asigna probabilidad $1/n$ a cada elemento de \mathbf{X} . Se llama muestra bootstrap de tamaño n a una muestra de tamaño n obtenida a partir de \hat{F} . Se denota por

$$\mathbf{X}^* = (X_1^*, \dots, X_n^*) .$$

Obsérvese que \mathbf{X}^* no tiene por qué ser igual a \mathbf{X} .

En otras palabras, \mathbf{X}^* se obtiene muestreando **con reemplazamiento** sobre \mathbf{X} . Por eso se llama *remuestreo*, porque se toman muestras a partir de una primera muestra.

Ejemplo 6.1 Sea una muestra $\mathbf{X} = (2, 1, 1, 0)$. Posibles muestras bootstrap pueden ser $\mathbf{X}^* = (1, 1, 0, 0)$, $\mathbf{X}^* = (1, 2, 0, 0)$, etc.

Para cada muestra bootstrap \mathbf{X}^* se puede calcular un valor $\hat{\theta}^*$ del estimador $\hat{\theta}$.

Definición 6.2 Al valor $\hat{\theta}^*$ obtenido a partir de una muestra bootstrap \mathbf{X}^* se le llama replicación bootstrap del estimador $\hat{\theta}$.

Definición 6.3 Llamamos estimación bootstrap del error estándar de $\hat{\theta}$ a la estimación obtenida utilizando la distribución \hat{F} en lugar de la distribución desconocida F . Concretamente, se calcula como $se(\hat{\theta}^*)$. Es decir, como el error estándar de $\hat{\theta}$ para muestras de tamaño n obtenidas a partir de \hat{F} .

Veamos en forma de algoritmo el procedimiento detallado de estimación bootstrap del error estándar.

ALGORITMO BOOTSTRAP

1. Sea \mathbf{X} una muestra de tamaño n .
2. Seleccionar B muestras bootstrap independientes

$$\mathbf{X}^{*1}, \dots, \mathbf{X}^{*B},$$

cada una de ellas obtenidas muestreando con reemplazamiento n elementos de la muestra original \mathbf{X} .

3. Evaluar las replicaciones bootstrap para cada una de las nuevas muestras. Sean $\hat{\theta}^*(b)$, $b = 1, \dots, B$ dichas replicaciones.
4. Estimar el error estándar $se(\hat{\theta})$ como la desviación típica muestral de las B replicaciones:

$$s\hat{e}_B(\hat{\theta}) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot) \right)^2},$$

donde

$$\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b).$$

Ejemplo 6.2 Sea una muestra $\mathbf{X} = (1, 2, 3, 4)$. Vamos a estimar el error estándar de la media usando 3 muestras bootstrap. Para ello calculamos la distribución empírica, obteniendo la siguiente:

$$\hat{F}(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{4} & \text{si } 1 \leq x < 2 \\ \frac{2}{4} = \frac{1}{2} & \text{si } 2 \leq x < 3 \\ \frac{3}{4} & \text{si } 3 \leq x < 4 \\ \frac{4}{4} = 1 & \text{si } x \geq 4 \end{cases}$$

es decir,

$$\hat{F}(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{\lfloor x \rfloor}{4} & \text{si } 1 \leq x < 4 \\ 1 & \text{si } x \geq 4 \end{cases}$$

Ahora necesitamos números aleatorios para obtener las muestras a partir de \hat{F} . Usaremos, por ejemplo, el siguiente generador:

$$x_{n+1} = (x_n + 7) \pmod{10} .$$

Se requieren 12 números para obtener las dos muestras. Si empezamos por el 0, esos números serán 0, 7, 4, 1, 8, 5, 2, 9, 6, 3, 0, 7 que llevados al intervalo $[0, 1)$ son 0, 0.7, 0.4, 0.1, 0.8, 0.5, 0.2, 0.9, 0.6, 0.3, 0 y 0.7.

A partir de estos números, aplicamos el método de inversión a \hat{F} y obtenemos las siguientes dos muestras:

$$\mathbf{X}^{*1} = (1, 3, 2, 1)$$

$$\mathbf{X}^{*2} = (4, 3, 1, 4)$$

$$\mathbf{X}^{*3} = (3, 1, 1, 3)$$

Las correspondientes replicaciones bootstrap son

$$\hat{\theta}^*(1) = \frac{7}{4} = 1.75$$

$$\hat{\theta}^*(2) = \frac{12}{4} = 3$$

$$\hat{\theta}^*(3) = \frac{8}{4} = 2$$

con lo que

$$\hat{\theta}^*(\cdot) = \frac{1.75 + 3 + 2}{3} = 2.25$$

y por tanto

$$\hat{s}_B(\hat{\theta}) = \sqrt{\frac{1}{2}((1.75 - 2.25)^2 + (3 - 2.25)^2 + (2 - 2.25)^2)} = 0.66 \text{ ,}$$

que es el valor estimado del error estándar.

6.3 Jackknife

A continuación, presentamos una técnica similar a la anterior, pero que difiere en el proceso de remuestreo.

Definición 6.4 Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra de tamaño n . Se llama muestra jackknife a cada una de las n muestras de tamaño $n - 1$ que se pueden obtener a partir de \mathbf{X} quitando un elemento cada vez:

$$\mathbf{X}_{(i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \quad i = 1, \dots, n \text{ .}$$

Definición 6.5 Se define la estimación jackknife del error estándar de un estimador $\hat{\theta}$ como

$$\hat{s}e_{\text{jack}} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n \left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} \right)^2},$$

donde $\hat{\theta}_{(i)}$ es la replicación de $\hat{\theta}$ para la muestra $\mathbf{X}_{(i)}$ y

$$\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}.$$

Veamos en forma de algoritmo el procedimiento detallado de estimación jackknife del error estándar.

ALGORITMO JACKKNIFE

1. Sea \mathbf{X} una muestra de tamaño n .
2. Sean las n muestras jackknife

$$\mathbf{X}_{(1)}, \dots, \mathbf{X}_{(n)}.$$

3. Evaluar las replicaciones jackknife para cada una de las nuevas muestras. Sean $\hat{\theta}_{(i)}$, $i = 1, \dots, n$ dichas replicaciones.
4. Estimar el error estándar $\hat{s}e_{\text{jack}}(\hat{\theta})$ como :

$$\hat{s}e_{\text{jack}} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n \left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} \right)^2},$$

donde $\hat{\theta}_{(i)}$ es la replicación de $\hat{\theta}$ para la muestra $\mathbf{X}_{(i)}$ y

$$\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}.$$

Ejemplo 6.3 Vamos a obtener mediante la técnica jackknife una estimación del error estándar de la media para la misma muestra del ejemplo 6.2. Para ello obtenemos las 4 posibles muestras jackknife que se pueden obtener a partir de \mathbf{X} , que son:

$$\mathbf{X}_{(1)} = (2, 3, 4)$$

$$\mathbf{X}_{(2)} = (1, 3, 4)$$

$$\mathbf{X}_{(3)} = (1, 2, 4)$$

$$\mathbf{X}_{(4)} = (1, 2, 3)$$

Ahora calculamos las correspondientes réplicas:

$$\hat{\theta}_{(1)} = \frac{2 + 3 + 4}{3} = 3$$

$$\hat{\theta}_{(2)} = \frac{1 + 3 + 4}{3} = 2.67$$

$$\hat{\theta}_{(3)} = \frac{1 + 2 + 4}{3} = 2.33$$

$$\hat{\theta}_{(4)} = \frac{1 + 2 + 3}{3} = 2$$

con lo que

$$\hat{\theta}_{(\cdot)} = \frac{3 + 2.67 + 2.33 + 2}{4} = 2.5 \text{ .}$$

En definitiva, la estimación jackknife del error estándar es

$$\widehat{\text{se}}_{\text{jack}} = \sqrt{\frac{3}{4}((3 - 2.5)^2 + (2.67 - 2.5)^2 + (2.33 - 2.5)^2 + (2 - 2.5)^2)} = 0.65 \text{ .}$$

Bibliografía

- [1] B. EFRON, R.J. TIBSHIRANI (1993). *An introduction to the bootstrap*. Chapman & Hall.
- [2] G.S. FISHMAN (1996). *Monte Carlo. Concepts, algorithms and applications*. Springer.
- [3] J.E. GENTLE (1998). *Random number generation and Monte Carlo methods*. Springer.
- [4] W.J. KENNEDY, J.E. GENTLE (1980). *Statistical computing*. Marcel Dekker Inc.
- [5] J. KLEIJNEN, W.V. GROENENDAAL (1992). *Simulation: a statistical perspective*. Wiley.
- [6] D.E. KNUTH (1986). *El arte de programar ordenadores*. Reverté.
- [7] B.J.T. MORGAN (1995). *Elements of simulation*. Chapman & Hall.
- [8] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, B.P. FLANNERY (1992). *Numerical recipes in C. The art of scientific computing. Second edition*. Cambridge University Press.
- [9] D. RÍOS, S. RÍOS, J. MARTÍN (1997). *Simulación. Métodos y aplicaciones*. Ra-Ma.
- [10] R.Y. RUBINSTEIN (1981). *Simulation and the Monte Carlo method*. Wiley.
- [11] R.Y. RUBINSTEIN, B. MELAMED (1998). *Modern simulation and modeling*. Wiley.

