

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE CIENCIAS



PRÁCTICAS PRE PROFESIONALES
SIMULACIÓN ESTADÍSTICA Y EL
MÉTODO DE MONTE CARLO
CON CADENAS DE MARKOV

AUTOR:

RABI PAOLO CAMARENA HUAMAN

ASESOR:

MG. WILLIAM ECHEGARAY CASTILLO

LIMA-PERÚ

2021

*A mi mamá Ana Maria Huaman Cenizario, quien siempre me apoyo en seguir
mis estudios en la UNI.*

Resumen

En este trabajo mostramos una breve introducción a la teoría de la simulación, la cual permite establecer métodos alternativos para tratar una diversidad de problemas de modelamiento matemático desde un enfoque estadístico. En particular, describimos el método de Monte Carlo con cadenas de Markov y elaboramos una simulación para el modelo de Ising en dimensión 2.

Abstract

In this work we show a brief introduction to the simulation theory, which allows to establish alternative methods to treat a variety of mathematical modeling problems from a statistical approach. In particular, we describe the Monte Carlo method with Markov chains and develop a simulation for the Ising model in dimension 2.

Índice general

Resumen	1
Índice General	2
Introducción	3
Justificación	3
Objetivos del trabajo	4
Marco Teórico	4
1. Generación de números y variables aleatorias	6
1.1. Espacios de probabilidad y variables aleatorias	6
1.2. Generación de números aleatorios	12
1.3. Generación de variables aleatorias	16
1.3.1. El método de la transformada inversa	16
1.3.2. El método de aceptación y rechazo	22
2. Análisis estadístico y datos simulados	28
2.1. Análisis estadístico de datos simulados	28
2.2. EL Teorema central del límite	29
2.3. Estimación puntual	33
2.4. Estimación por intervalos de confianza	38
2.5. Método por Monte Carlo	42

3. Método de Monte Carlo con Cadenas de Markov	48
3.1. Cadenas de Markov	49
3.2. El algoritmo de Hastings-Metropolis	52
3.2.1. El muestreador de Gibbs	54
3.3. El modelo de Ising	56
3.3.1. Transición de fase	57
3.3.2. Simulación de la medida de Gibbs	58
Conclusiones	61
Apéndice	62

Introducción

Justificación

Es conocido que si una función f es continua y acotada en un intervalo $[a, b]$ entonces existe la integral de dicha función en ese intervalo. También es sabido que el cálculo del valor de esta integral se realiza hallando la antiderivada, es decir, hallando una función F definida en el intervalo $[a, b]$ de modo que se cumpla que $F' = f$ en dicho intervalo. Sin embargo, en muchos casos no es posible, mediante algún método de cálculo elemental, encontrar la respectiva antiderivada. Asimismo, es común que en estos casos precisemos solamente de una aproximación del valor de la integral. Una de las formas de abordar este problema es vía métodos numéricos, aquí se presenta otra forma de conseguir ello, el llamado método de integración por Monte Carlo.

Tal método consiste en considerar la integral como un parámetro de una población de variables aleatorias, más específicamente la media, así al tomar una muestra podemos estimar dicho parámetro (el valor de la integral) como la media de la muestra tomada. La ley de los grandes números y teorema central del límite nos permiten garantizar que en efecto este procedimiento nos lleva a una aproximación del valor de la integral (cuando el tamaño de la muestra es lo suficientemente grande) y dar un intervalo donde se puede estar casi seguro de encontrar el valor exacto de la integral.

En general, la teoría de simulación nos da un enfoque alterno de un problema de modelamiento matemático, y esto se refleja en los métodos que desarrolla para resolverlo, obtener la información pertinente. Es importante mencionar que la simulación se apoya en el poder computacional, que ha ido en crecimiento en las últimas décadas, y por lo tanto ha ido en crecimiento a la par con este. Así, este trabajo monográfico busca dar un primer acercamiento a la teoría de la simulación, la cual es usada a menudo.

Objetivos del trabajo

El objetivo principal del presente trabajo es realizar una breve introducción a la teoría de la simulación, la cual permite establecer métodos alternativos para tratar una diversidad de problemas de modelamiento matemático desde un enfoque estadístico. En particular, describir el método de Monte Carlo con cadenas de Markov y elaborar una simulación para el modelo de Ising en dimensión 2.

Marco teórico

El presente trabajo sigue el desarrollo realizado en el libro [6], y consta de tres capítulos y un apéndice. En el capítulo 1 se hace uso del poder del cálculo computacional para generar números con un comportamiento aleatorio. En seguida se muestra como a partir de la generación de números aleatorios se pueden generar valores de una variable aleatoria discreta o continua. En el capítulo 2 se establecen las técnicas estadísticas que permitirán estimar un parámetro característico de una población de variables aleatorias mediante el estudio de una muestra. También, se desarrolla el método de integración por Monte Carlo como un caso particular del problema de estimación ya estudiado. En el capítulo 3 se muestra una metodología general que es muy práctica al momento de simular una distribución arbitraria, conocida como método de Monte Carlo con cadenas de Markov así como se presenta un ejemplo cercano a este tipo de metodologías como es el modelo de Ising. Especialmente se trata el problema de transición de fase para el modelo de Ising en dimensión

$$d = 2.$$

Capítulo 1

Generación de números y variables aleatorias

En este capítulo se hace uso del poder del cálculo computacional para generar números con un comportamiento aleatorio. En seguida se muestra como a partir de la generación de números aleatorios se pueden generar valores de una variable aleatoria discreta o continua, veamos algunos conceptos previos.

1.1. Espacios de probabilidad y variables aleatorias

La teoría aquí expuesta es la justa y necesaria para nuestros fines, demostrar el Teorema Central del Límite. El libro seguido para esta sección es el referenciado en [2].

Probabilidad y condicionamiento

Definición 1. *Un σ -álgebra es una colección $\mathcal{F} \subset \mathcal{P}(\Omega)$ de subconjuntos del espacio muestral estable (cerrada) bajo uniones contables, al tomar complemento, y que contiene al evento vacío \emptyset . La dupla (Ω, \mathcal{F}) se denomina espacio medible. Un evento es un elemento de \mathcal{F} .*

$\mathcal{P}(\Omega)$ designa a la colección de todos los subconjuntos de Ω ; \mathcal{F} es una colección de

subconjuntos de Ω . La estabilidad bajo unión contable es escrita formalmente: para toda sucesión de eventos $\{E_n: n \geq 1\} \subset \mathcal{F}$, se cumple que la unión es un evento, $\bigcup_{n \geq 1} E_n \in \mathcal{F}$. La estabilidad al tomar complemento es escrita formalmente: el complemento $E^c := \Omega \setminus E$ de todo evento $E \in \mathcal{F}$ es un evento, $E^c \in \mathcal{F}$.

Observación 1.

1. Si Ω es finito o contable, tomamos por defecto al σ -álgebra $\mathcal{P}(\Omega)$.
2. $E \cap F = (E^c \cup F^c)^c \in \mathcal{F}$ para todo par de eventos $E, F \in \mathcal{F}$.
3. $E \setminus F = E \cap F^c \in \mathcal{F}$ para todo par de eventos $E, F \in \mathcal{F}$.

Definición 2. Una probabilidad sobre un espacio medible (Ω, \mathcal{F}) es una función $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$ que verifica $\mathbb{P}(\Omega) = 1$, y la propiedad de aditividad contable:

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n)$$

para toda sucesión $\{A_n: n \geq 1\}$ de eventos disjuntos dos a dos. La terna $(\Omega, \mathcal{F}, \mathbb{P})$ se denomina espacio de probabilidad. Un evento de probabilidad nula se llama evento insignificante. Un evento se dice casi seguro si su complemento es insignificante.

Proposición 1 (Propiedades de una probabilidad). Toda probabilidad \mathbb{P} cumple las siguientes propiedades:

1. El evento imposible es insignificante, es decir, $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mathbb{P}(A_n)$ para toda sucesión $\{A_n: n \geq 1\}$ de eventos.
3. \mathbb{P} es creciente: para todo par de eventos A, B con $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.
4. $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$ para todo par de eventos A, B tales que $A \subset B$. En particular $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ para todo evento A .

5. Toda intersección contable de eventos casi seguros es casi seguro, es decir, si

$\{A_n: n \geq 1\}$ es una sucesión de eventos con $\mathbb{P}(A_n) = 1$ para todo $n \geq 1$, entonces

$$\mathbb{P}\left(\bigcap_{n \geq 1} A_n\right) = 1.$$

6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ para todo par de eventos A, B .

7. \mathbb{P} es continua:

$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ para toda sucesión creciente de eventos $\{A_n: n \geq 1\}$

(es decir, $A_n \subset A_{n+1}$); así como $\mathbb{P}\left(\bigcap_{n \geq 1} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n)$ para toda sucesión decreciente de eventos $\{A_n: n \geq 1\}$ (es decir, $B_{n+1} \subset B_n$).

Ejemplo 1. Probabilidad discreta sobre un conjunto numerable $\Omega = \{w_j\}_{j \geq 1}$: Se define por la lista (sucesión) de probabilidades de los singletons $p_j = \mathbb{P}(\{w_j\})$, tal que $\sum_{j \geq 1} p_j = 1$.

De hecho, para todo evento $A \subset \Omega$ tenemos

$$\mathbb{P}(A) := \sum_{w_j \in A} p_j$$

Definición 3 (Probabilidad condicional). Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y el evento $C \in \mathcal{F}$, no insignificante. La probabilidad condicional relativa (o “dado C ”) a C esta definida por:

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}.$$

Se verifica inmediatamente que en efecto $\mathbb{P}(\cdot|C)$ es una probabilidad sobre (Ω, \mathcal{F}) .

Definición 4. Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se dice que dos eventos A y B son independientes cuando $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Podemos redefinir la independencia usando el concepto de probabilidad condicional. Sea B un evento no insignificante entonces A y B son eventos *independientes* si y solo si

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

Definición 5. Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se dice que los eventos A_1, A_2, \dots, A_n son independientes cuando

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_k})$$

para todos $1 \leq i_1 \leq \dots \leq i_k \leq n$. Una sucesión de eventos se dice independiente si toda subsucesión finita esta constituida de eventos independientes.

Variables Aleatorias

Definición 6 (Variable aleatoria). Una variable aleatoria, abreviado v.a., es una función V definida sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con valores sobre \mathbb{R}^d , tal que $\{V \in E\} := V^{-1}(E) \in \mathcal{F}$, para todo E abierto o cerrado de \mathbb{R}^d . La ley de V es la probabilidad $P_V := \mathbb{P} \circ V^{-1}$ definida sobre el espacio medible $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ como sigue:

$$P_V(E) := \mathbb{P}(V^{-1}(E)) = \mathbb{P}(V \in E),$$

en este caso, es usual denotar $V \sim P_V$.

Si $d = 1$, se trata de una variable aleatoria real, abreviado v.a.r. El σ -álgebra generada por las n variables aleatorias V_1, \dots, V_n esta constituido por el σ -álgebra $\sigma\{V_1, \dots, V_n\} \subset \mathcal{F}$ que contiene todos los eventos de la forma $(V_1, \dots, V_n)^{-1}(E)$ donde E es un abierto o cerrado en \mathbb{R}^d .

Una función de Ω sobre \mathbb{R}^d es una v.a. si y sólo si sus coordenadas (sin importar la base de \mathbb{R}^d) son v.a.r. Una combinación lineal de v.a. es también una v.a. Un producto de v.a.r. es una v.a.r. La composición de una v.a. sobre \mathbb{R}^d por una función continua de \mathbb{R}^d sobre $\mathbb{R}^{d'}$ es una v.a. sobre $\mathbb{R}^{d'}$.

Observación 2. La función de distribución de una v.a. real X es la función $F_X: \mathbb{R} \rightarrow [0, 1]$ definida por la regla

$$F_X(x) := \mathbb{P}(X \leq x) = P_X([-\infty, x]), \quad x \in \mathbb{R},$$

que es creciente, continua a la derecha, y verifica $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$. Ella determina la ley P_X de X . Se llama mediana de X (o de su ley) a cualquier m tal que

$$F_X(m^-) := \lim_{x \rightarrow m^-} F_X(x) \leq \frac{1}{2} \leq F_X(m).$$

Definición 7. La esperanza de una variable aleatoria V definida sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ es la integral respecto de la probabilidad \mathbb{P} :

$$\mathbb{E}[V] = \int_{\Omega} V d\mathbb{P} \quad (1.1)$$

Ella esta bien definida cuando $\mathbb{E}[|V|]$ es finita, en este caso se dice que V es integrable.

Sea $A \subset \Omega$ no vacío, se denota por 1_A a la función indicadora de A la cual esta definida por

$$1_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in \Omega \setminus A. \end{cases}$$

Definición 8 (Variable aleatoria discreta). Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Una variable aleatoria $V: \Omega \rightarrow \mathbb{R}^d$ se llama discreta si $V(\Omega)$ es numerable. Toda variable discreta puede escribirse de la siguiente forma

$$V = \sum_{j \geq 1} v_j 1_{\{V=v_j\}},$$

donde $\{[V = v_j]\}_{j \geq 1} \subset \Omega$ es una sucesión de eventos y $\{v_j\}_{j \geq 1}$ es una sucesión en \mathbb{R}^d de vectores distintos dos a dos.

Si V es una variable aleatoria discreta, la esperanza de V se calcula con la fórmula:

$$\mathbb{E}[V] = \sum_{j \geq 1} v_j \mathbb{P}(V = v_j).$$

En este caso, se llama función de masa de probabilidad a la sucesión $\{p_j, j \geq 1\}$ definida por

$$p_j = \mathbb{P}(V = v_j), \quad j \geq 1.$$

Proposición 2. Sea V una v.a. sobre \mathbb{R}^d y f una función continua acotada de \mathbb{R}^d a $\mathbb{R}^{d'}$.

Entonces

$$\mathbb{E}[f \circ V] = \int_{\Omega} f \circ V d\mathbb{P} = \int_{\mathbb{R}^d} f dP_V$$

Cuando la variable V es discreta,

$$\mathbb{E}[f \circ V] = \sum_v f(v) \mathbb{P}(V = v)$$

cuya serie es convergente si $f \circ V$ es integrable.

La esperanza es lineal: para todas las variables aleatorias V_1, \dots, V_n integrables y todos los escalares $\lambda_1, \dots, \lambda_n$, se tiene

$$\mathbb{E} \left[\sum_{i=1}^n \lambda_i V_i \right] = \sum_{i=1}^n \lambda_i \mathbb{E}[V_i]$$

Definición 9. Una variable aleatoria V con valores en \mathbb{R}^d admite una densidad h si

$$\mathbb{E}[V] = \int_{\mathbb{R}^d} Id \cdot h d\lambda.$$

Una variable V que admite una densidad se dice absolutamente continua. Así, si f es una función continua y acotada de \mathbb{R}^d a $\mathbb{R}^{d'}$ entonces

$$\mathbb{E}[f \circ V] = \int_{\mathbb{R}^d} f \cdot h d\lambda.$$

Definición 10. Una variable aleatoria V se dice p -integrable ($1 \leq p < \infty$), denotado $V \in L_p$, cuando $\mathbb{E}[\|V\|^p] < \infty$. En particular, si $p = 1$ entonces V se dice integrable y si $p = 2$ entonces V se dice cuadrado integrable.

La varianza de una v.a. real V cuadrado integrable es

$$\text{Var}(V) = \mathbb{E}[|V - \mathbb{E}(V)|^2] = \mathbb{E}[V^2] - \mathbb{E}[V]^2.$$

Definición 11. Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Las v.a. V_1, \dots, V_n, \dots definidas sobre (Ω, \mathcal{F}) se llaman independientes cuando para todos los conjuntos B_1, \dots, B_n, \dots abiertos o cerrados en \mathbb{R}^d , los eventos $\{V_1 \in B_1\}, \dots, \{V_n \in B_n\}, \dots$ son independientes.

Proposición 3. Las v.a. V_1, \dots, V_n, \dots son independientes si y sólo si

$$\mathbb{E}[f_1(V_1) \times \dots \times f_n(V_n)] = \mathbb{E}[f_1(V_1)] \times \dots \times \mathbb{E}[f_n(V_n)]$$

para todo $n \geq 1$ y todas las funciones (medibles) positivas f_1, \dots, f_n .

Expresado de otra forma: la probabilidad inducida por (V_1, \dots, V_n, \dots) es el producto de probabilidades inducidas por cada V_n . Em particular, cuando las v.a. reales independientes V_j admiten funciones de densidad h_j sobre \mathbb{R} , entonces la v.a. (V_1, \dots, V_n) admite la densidad $h_1 \otimes \dots \otimes h_n$ sobre \mathbb{R}^n (definida por $h_1 \otimes \dots \otimes h_n(x_1, \dots, x_n) = h_1(x_1) \times \dots \times h_n(x_n)$).

Observación 3.

1. Los eventos A_1, \dots, A_n, \dots son independientes si y sólo si las v.a. $1_{A_1}, \dots, 1_{A_n}, \dots$ son independientes.
2. Si las v.a. $V_1, \dots, V_n, \dots \in L^2$ son independientes son independientes, entonces la covarianza de dos cualesquiera es nula. El recíproco es falso, un contraejemplo es: si las variables son independientes 2 a 2 pero no independientes.
3. Las v.a. discretas V_1, \dots, V_n, \dots son independientes si y sólo si para todos los valores v_1, \dots, v_n, \dots los eventos $\{V_1 = v_1\}, \dots, \{V_n = v_n\}, \dots$ son independientes.

Proposición 4. Sean V_1, \dots, V_n, \dots v.a. independientes con valores en \mathbb{R}^d y sea f_1, \dots, f_n, \dots una sucesión de funciones continuas de \mathbb{R}^d a $\mathbb{R}^{d'}$. Entonces las v.a. $f_1(V_1), \dots, f_n(V_n), \dots$ con valores en $\mathbb{R}^{d'}$, también son independientes.

1.2. Generación de números aleatorios

Lo central de un estudio de simulación consiste en obtener resultados de un experimento sin necesidad de realizarlo, ya que esto nos permite estudiar un fenómeno real basándose en el modelo planteado para el experimento. Por ello, para empezar un estudio de simulación

se requiere una buena capacidad de generación de números aleatorios, pues tomando como base esto se puede contruir un sistema o fenómeno aleatorio. En la práctica solo podemos generar números pseudoaleatorios, que no son sino números generados por algoritmos determinísticos que se “disfrazan” de aleatorios. De esta manera en el lenguaje de la simulación usamos los términos número aleatorio y número pseudoaleatorio sin distinguir entre ellos, salvo la situación lo amerite.

Un método muy común, llamado **método congruencial multiplicativo**, para generar números aleatorios consiste en comenzar con un valor inicial x_0 , llamada **semilla**, y luego calcular de manera recursiva los valores sucesivos x_n , $n \geq 1$, haciendo

$$x_n = (ax_{n-1}) \text{ mód } m, \quad (1.2)$$

donde a y m son enteros positivos dados. Los valores de x_n se obtienen como el residuo de dividir ax_{n-1} entre m , así $x_n = 0, 1, \dots, m-1$ y la cantidad x_n/m (llamado **número aleatorio** o, más precisamente, **número pseudoaleatorio**) se considera como una aproximación del valor de una variable aleatoria uniforme en $]0, 1[$.

Como cada uno de los números generados, x_n , toma uno de los valores $0, 1, \dots, m-1$, se tiene que después de un número finito (a lo más m) de valores generados, alguno debe repetirse, y, una vez que esto ocurre, los valores siguientes de la sucesión también se repiten. Así, queremos hallar las constantes a y m tales que satisfagan las tres condiciones siguientes:

1. Para cualquier semilla x_0 , la sucesión resultante x_n tiene la “apariencia” de ser una sucesión de variables aleatorias independientes y uniformes en $]0, 1[$.
2. Para cualquier semilla x_0 , la cantidad de variables que se puedan generar, antes de que comience la repetición, debe ser “grande”.
3. Los valores de las variables se pueden calcular de manera eficiente en una computadora.

Un criterio que parece ser útil para cumplir las tres condiciones anteriores es tomar m como

un número primo muy grande, adecuado al tamaño de la palabra de la computadora¹. Para una máquina con longitud de palabra igual a 32 (donde el primer bit es de signo), se ha probado que las elecciones de $m = 2^{31} - 1$ y $a = 7^5$ producen las propiedades deseadas. Para mayor detalle se insta a revisar el capítulo 3 del libro [1], si se requiere mayor profundidad esto se encuentra en los capítulos 1 y 2 del libro [4].

Una variante del método anterior, llamado **método congruencial mixto**, el cual esta basado en una ecuación recursiva de la forma

$$x_n = (ax_{n-1} + c) \mod m.$$

¹La palabra de una computadora es un estándar que sirve para transferir datos dentro de una computadora. Una palabra está conformada por un grupo de bits donde un bit es la unidad básica de información y toma solo dos valores, 0 y 1. La cantidad de bits que conforman a una palabra es conocida como la longitud de la palabra y está determinada por la arquitectura de la máquina.

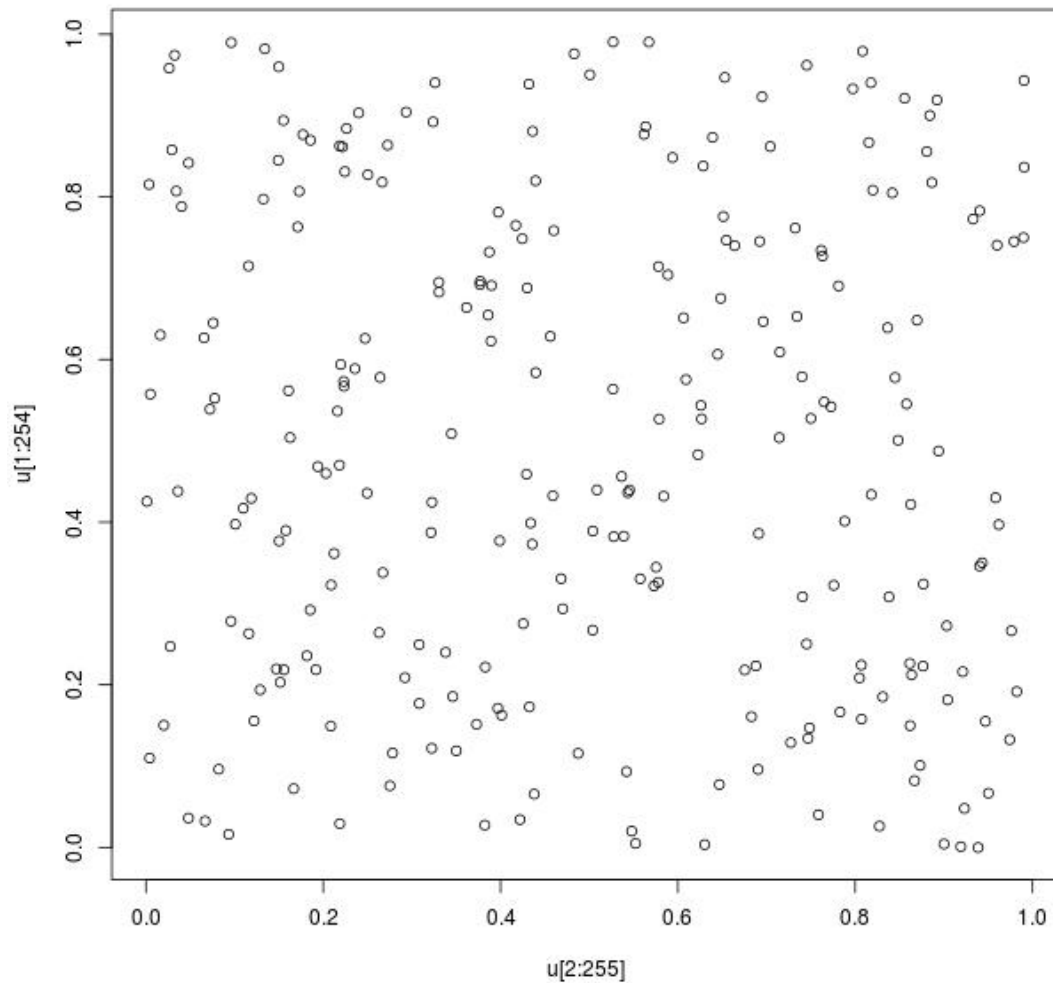


Figura 1.1: El método congruencial mixto

La mayor parte de los lenguajes de programación tienen por defecto una función generadora de números pseudoaleatorios integrados a ellos que se usa con el fin de generar tales números. Por ejemplo, el lenguaje R usa la instrucción $u = \text{runif}(1)$ para generar un número aleatorio, es decir, un valor según una distribución uniforme en $]0, 1[$.

Un ejemplo del funcionamiento del método congruencial mixto se muestra en el gráfico 1.1, para mostrar esta figura se programó en R un serie de sentencias, considerando

$a = 314159269$, $c = 453806245$ y $m = 2^{31} - 1$, que al ejecutarlas calculan una secuencia de números x_1, x_2, \dots, x_{256} . Note que con tales parámetros fijados esta secuencia de números tiene la apariencia de ser una sucesión de números aleatorios, lo cual se observa en la dispersión de los puntos abarcando toda el área cuadrada.

Como punto de partida en la simulación de sistemas por computadora, supondremos que podemos generar una sucesión de números pseudoaleatorios que se pueden considerar como una aproximación a los valores de una sucesión de variables aleatorias independientes y uniformes en $]0, 1[$. Es decir, no exploraremos las interesantes cuestiones teóricas relacionadas con la construcción de “buenos” generadores de números pseudoaleatorios, nuevamente se cita el libro [4] para abordar dicha cuestión. Así, en adelante, se asumirá que se tiene una “caja negra” que proporciona un número aleatorio si así se solicita.

1.3. Generación de variables aleatorias

1.3.1. El método de la transformada inversa

El método de la transformada (o transformación) inversa, también conocido como método de la inversa de la transformada, es un método para la generación de números aleatorios de cualquier distribución de probabilidad cuando se conoce la inversa de su función de distribución. Este método es en general aplicable, pero puede resultar muy complicado obtener una expresión analítica de la inversa para algunas distribuciones de probabilidad. veamos para el caso discreto

Suponga que queremos generar el valor de una variable aleatoria discreta X con función de masa de probabilidad²

$$\mathbb{P}(X = x_j) = p_j, \quad j = 0, 1, \dots, \sum_j p_j = 1. \quad (1.3)$$

²Ver definición 8.

Para esto, generamos un número aleatorio U ; y, como U es el valor de una variable aleatoria distribuida uniformemente en $]0, 1[$, se asigna

$$X = x_j \quad \text{si} \quad \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i, \quad j = 0, 1, 2, \dots \quad (1.4)$$

Entonces, para cada $j = 0, 1, 2, \dots$ se cumple

$$\begin{aligned} \mathbb{P}(X = x_j) &= \mathbb{P}\left(\sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i\right) \\ &= \sum_{i=0}^j p_i - \sum_{i=0}^{j-1} p_i \\ &= p_j, \end{aligned}$$

y por lo tanto, X tiene la distribución deseada.

Observación 4. Si los valores x_i , $i \geq 0$, están ordenados de modo que $x_0 < x_1 < x_2 < \dots$ y F denota la función de distribución de X , entonces $F(x_j) = \sum_{i=0}^j p_i$, $j \geq 0$, la ecuación (1.4) se convierte en

$$X = x_j, \quad \text{si} \quad F(x_{j-1}) \leq U < F(x_j). \quad (1.5)$$

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

El método de la transformada inversa puede escribirse en forma algorítmica de la siguiente manera

Algoritmo 1 (Transformada inversa en el caso discreto).

1. Generar un número aleatorio U .
2. $j = 0$, $x = x_0$, $F = 0$.

3. Si $U < F$ hacer $X = x$ y terminar. Sino, hacer $x = x_{j+1}$, $F = F + p_{j+1}$, $j = j + 1$.

4. Ir al paso 3.

Observación 5. El tiempo necesario para generar una variable aleatoria discreta mediante este método es proporcional al número de intervalos en los que se realizan las búsquedas. Por esta razón, a veces vale la pena considerar los valores posibles x_j de X en orden decreciente de los p_j .

Ejemplo 2 (Generación de una variable aleatoria geométrica). Sea X una variable aleatoria geométrica con parámetro p , es decir, su función de masa de probabilidad es

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

Además, consideraremos que X representa el tiempo en el ocurre el primer éxito, cuando se realizan ensayos independientes, de los que cada uno es un éxito con probabilidad p . Como

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

podemos generar el valor de X al generar un número aleatorio U y hacer X igual al valor j para el que se verifica las siguientes desigualdades, dadas en (1.4),

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

o, en forma equivalente,

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

como la función logaritmo natural es una función estrictamente creciente, lo anterior también equivale a

$$j \log(q) < \log(1 - U) \leq (j - 1) \log(q).$$

Es decir, podemos definir X como

$$X = j \geq 1 \text{ tal que } j - 1 \leq \frac{\log(1 - U)}{\log(q)} < j.$$

Luego, denotando la función máximo entero³ como $\text{Ent}(\cdot)$, podemos expresar X como

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

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

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

también es geométrica con parámetro p .

Como ya vimos en la sección anterior asumiremos que tenemos a la mano un generador de números aleatorios, y ahora, nos proponemos a partir de ello generar valores de una variable aleatoria discreta. A continuación se describen unos algunos métodos desarrollados para tal fin. veamos el caso continuo

El método de la transformada inversa para variables aleatorias discretas tiene su análogo en el caso de variables aleatorias continuas. Por ello generalizamos este método para considerar ambos casos. Así se obtiene el algoritmo de la transformada inversa, este se basa en el siguiente teorema, previamente se ve un lema previo.

Lema 1. Sea F una función de distribución y sea la función F^{-1} definida como sigue

$$\begin{aligned} F^{-1} :]0, 1[&\rightarrow \mathbb{R} \\ u &\mapsto F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}. \end{aligned}$$

Entonces F^{-1} tiene las propiedades siguientes: es no decreciente, continua por la izquierda y además, para todo $u \in]0, 1[$ y $x \in \mathbb{R}$ se cumple

$$F^{-1}(u) \leq x \Leftrightarrow u \leq F(x).$$

³Sea $x \in \mathbb{R}$, la función máximo entero en x queda determinado como el entero $\text{Ent}(x)$ que verifica las siguientes desigualdades: $\text{Ent}(x) \leq x < \text{Ent}(x) + 1$.

Demostración. Sea F una función de distribución y F^{-1} definida como en el enunciado del lema. Veamos que F^{-1} es no decreciente, en efecto, si $t \leq u$ en $]0, 1[$ entonces $\{x \in \mathbb{R} : F(x) \geq u\} \subset \{x \in \mathbb{R} : F(x) \geq t\}$ y por tanto

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\} \leq \inf\{x \in \mathbb{R} : F(x) \geq u\} = F^{-1}(u).$$

Además, F^{-1} es continua por la izquierda, pues para todo $x \in \mathbb{R}$ tenemos que

$$(F^{-1})^{-1}(]-\infty, x]) =]0, F(x)[\quad \text{es abierto en }]0, 1[.$$

Probamos la propiedad de F^{-1} , sean $u \in]0, 1[$ y $x \in \mathbb{R}$, entonces, $F^{-1}(u) \leq x$ implica $x \in \{y \in \mathbb{R} : F(y) \geq u\}$, es decir, $u \leq F(x)$, y de manera similar $u \leq F(x)$ implica que $x \in \{y \in \mathbb{R} : F(y) \geq u\}$, entonces, $\inf\{y \in \mathbb{R} : F(y) \geq u\} \leq x$, es decir, $F^{-1}(u) \leq x$. \square

Teorema 1. Sea U una variable aleatoria uniforme en $]0, 1[$. Para cualquier función de distribución continua e invertible, F , la variable aleatoria X definida como

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

tiene distribución F , donde F^{-1} está definida en el lema 1.

Demostración. Sea U una variable aleatoria uniforme entonces la composición $X = F^{-1}(U)$ es también una variable aleatoria. Sea F_X la función de distribución de X , entonces para cada $x \in \mathbb{R}$ se tiene

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x).$$

Ahora, como F es una función de distribución, por la última propiedad de F^{-1} enunciada en el lema 1 se tiene que

$$\begin{aligned} F_X(x) &= \mathbb{P}(U \leq F(x)) \\ &= F(x), \end{aligned}$$

para todo $x \in \mathbb{R}$. Por lo tanto $F_X = F$. \square

Podemos representar el teorema anterior como un algoritmo que sigue la siguiente lógica:

Algoritmo 2 (El método de la transformada inversa).

1. Generar un número aleatorio U .

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

Ejemplo 3. Sea X una variable aleatoria exponencial con razón 1, entonces su función de distribución está dada por

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

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

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

de donde

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

Por lo tanto, para generar X generamos un número aleatorio U y hacemos

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

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

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

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

1.3.2. El método de aceptación y rechazo

Supóngase que tenemos un método para generar una variable aleatoria con función de densidad $g(x)$. Podemos usar este método como base para generar un valor a partir de la variable continua con función de densidad $f(x)$: primero se genera Y a partir de g y luego se acepta este valor generado con una probabilidad proporcional a $\frac{f(Y)}{g(Y)}$.

En concreto, sea c una constante tal que

$$\frac{f(y)}{g(y)} \leq c, \quad \forall y \in \mathbb{R}. \quad (1.6)$$

El algoritmo del rechazo es igual al del caso de variables aleatorias discretas, con la única diferencia de que las densidades reemplazan a las funciones de masa, como se muestra a continuación:

Algoritmo 3 (El método del rechazo).

1. Generar Y con densidad g .
2. Generar un número aleatorio U .
3. Si $U \leq \frac{f(Y)}{c \cdot g(Y)}$, hacer $X = Y$ y terminar. En caso contrario, regresar al paso 1.

Teorema 2. Sean f y g dos densidades de probabilidad y sea $c > 0$ de modo que se cumple la desigualdad (1.6). El algoritmo de rechazo genera una variable aleatoria X con densidad f , es decir, con función de distribución

$$F(x) = \int_{-\infty}^x f(t) dt, \quad \forall x \in \mathbb{R}.$$

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

Observación 6. Se tienen las siguientes observaciones:

1. Se debe observar que la forma en que “aceptamos el valor de Y con probabilidad $\frac{f(Y)}{c \cdot g(Y)}$ ” consiste en generar un número aleatorio U y luego aceptar Y si $U \leq \frac{f(Y)}{c \cdot g(Y)}$.
2. De la ecuación (1.6) se puede mostrar que $c \geq 1$ y que este c a su vez puede calcularse como

$$c = \sup_Y \frac{f(Y)}{g(Y)}.$$

3. El número de iteraciones del algoritmo necesarias, en promedio, para obtener un valor aceptado es c . Entonces, mientras más cerca esté c de 1 más eficaz será el método.

Ejemplo 4. Sea una variable aleatoria con la densidad gamma $(\frac{3}{2}, 1)$

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

donde $K = 1/\Gamma(\frac{3}{2}) = \frac{2}{\sqrt{\pi}}$. Esta variable aleatoria está concentrada en el eje positivo y tiene media $\frac{3}{2}$, por lo cual intentaremos la técnica de rechazo con una variable aleatoria exponencial con la misma media, la cual tiene una distribución que cumple esta condición. Por lo tanto, sea

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

Ahora, podemos hallar c al maximizar el cociente

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

Luego, derivando e igualando a cero la derivada resultante se obtiene que el valor máximo se alcanza cuando $x = \frac{3}{2}$. Por tanto,

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

y así

$$\frac{f(x)}{c \cdot g(x)} = \left(\frac{2ex}{3}\right)^{1/2} e^{-x/3}.$$

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

1. Generar un número aleatorio U_1 y hacer $Y = -\frac{3}{2} \log U_1$.
2. Generar un número aleatorio U_2 .
3. Si $U_2 < \left(\frac{2eY}{3}\right)^{1/2} e^{-Y/3}$, hacer $X = Y$. En caso contrario, regresar al paso 1.

El número promedio de iteraciones necesarias es

$$c = \frac{3^{3/2}}{(2\pi e)^{1/2}} = 1,257.$$

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

Específicamente, sea c una constante tal que

$$\frac{p_j}{q_j} \leq c \quad \forall j \geq 0 \text{ con } p_j > 0. \quad (1.7)$$

Ahora tenemos la siguiente técnica, llamada **método de rechazo** o **método de aceptación y rechazo**, para simular una variable aleatoria X con función de masa $p_j = \mathbb{P}(X = j)$. A continuación se escribe el método del rechazo en forma de algoritmo.

Algoritmo 4 (El método del rechazo).

1. Simular el valor de Y , con función de la masa de probabilidad q_j .
2. Generar un número aleatorio U .
3. Si $U \leq \frac{p_Y}{c \cdot q_Y}$, hacer $X = Y$ y terminar. En caso contrario, regresar al paso 1.

Teorema 3. Sean $\{p_j, \geq 0\}$ y $\{q_j, \geq 0\}$ dos de masa de probabilidad y sea $c > 0$ de modo que se cumple la desigualdad (1.7). El algoritmo de rechazo genera una variable aleatoria X tal que

$$\mathbb{P}(X = j) = p_j, \quad j = 0, 1, \dots$$

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

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

$$\begin{aligned} \mathbb{P}(Y = j, \text{aceptada}) &= \mathbb{P}(Y = j) \cdot \mathbb{P}(\text{aceptada} | Y = j) \\ &= q_j \frac{p_j}{c \cdot q_j} \\ &= \frac{p_j}{c}. \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}(\text{aceptada}) &= \mathbb{P}\left(\bigcup_j \{Y = j, \text{aceptada}\}\right) \\ &= \sum_j \mathbb{P}(Y = j, \text{aceptada}) \\ &= \sum_j \frac{p_j}{c} \\ &= \frac{1}{c}. \end{aligned}$$

Como cada iteración produce de manera independiente un valor aceptado con probabilidad $\frac{1}{c}$, vemos que el número de iteraciones necesarias sigue una distribución geométrica

con media c . Además,

$$\begin{aligned}
\mathbb{P}(X = j) &= \sum_n \mathbb{P}(j \text{ aceptada en la iteración } n) \\
&= \sum_n \mathbb{P}(\text{se requieren } n \text{ iteraciones, el valor aceptado es } j) \\
&= \sum_n \mathbb{P}(\text{se requieren } n \text{ iteraciones}) \cdot \mathbb{P}(\text{el valor aceptado es } j) \\
&= \sum_n \left(1 - \frac{1}{c}\right)^{n-1} \cdot \frac{p_j}{c} \\
&= p_j.
\end{aligned}$$

□

Observación 7.

1. Se debe observar que la forma en que “aceptamos el valor de Y con probabilidad $\frac{p_Y}{c \cdot q_Y}$ ” consiste en generar un número aleatorio U y luego aceptar Y si $U \leq \frac{p_Y}{c \cdot q_Y}$.
2. De la ecuación (1.7) se puede mostrar que $c \geq 1$ y que este c a su vez puede calcularse como

$$c = \sup_{j \geq 0} \frac{p_j}{q_j}.$$

3. El número de iteraciones del algoritmo necesarias para obtener un valor aceptado tiene una distribución geométrica con parámetro $1/c$ y por tanto tiene media c . Entonces, mientras más cerca esté c de 1 más eficaz será el método; esto es claro intuitivamente, pues mientras más cerca esté c de 1, más parecidas serán las dos funciones de masa $\{p_j\}$ y $\{q_j\}$.

Ejemplo 5. Suponga que queremos simular el valor de una variable aleatoria X que toma uno de los valores $1, 2, \dots, 10$ con probabilidades $0,11, 0,12, 0,10, 0,09, 0,08, 0,12, 0,10, 0,09, 0,10, 0,10$. Una posibilidad es utilizar el algoritmo de la transformada inversa, pero un mejor planteamiento consiste en usar el método de rechazo, con q como la densidad

uniforme discreta en $1, \dots, 10$. Es decir, $q_j = 0,10$, $j = 1, \dots, 10$. Para esta elección de $\{q_j\}$, podemos elegir c como

$$c = \max_j \frac{p_j}{q_j} = 1,2$$

de modo que el algoritmo sería el siguiente:

1. Generar un número aleatorio U_1 y hacer $Y = \text{Ent}(10U_1) + 1$
2. Generar un segundo número aleatorio U_2 .
3. Si $U_2 \leq \frac{p_Y}{0,12}$, hacer $X = Y$ y terminar. En caso contrario, regresar al paso 1.

La constante 0,12 del paso 3 se debe a que $c \cdot q_Y = \frac{1,2}{10} = 0,12$. En promedio, este algoritmo requiere solo 1,2 iteraciones para obtener el valor generado de X .

Capítulo 2

Análisis estadístico y datos simulados

En este capítulo se establecen las técnicas estadísticas que permitirán estimar un parámetro característico de una población de variables aleatorias mediante el estudio de una muestra. También, se desarrolla el teorema central del límite y finalmente el método de integración por Monte Carlo como un caso particular del problema de estimación.

2.1. Análisis estadístico de datos simulados

Por lo general, los estudios de simulación se realizan para determinar el valor de cierta cantidad θ relacionada con un modelo estocástico¹ particular. Una simulación del sistema² en cuestión produce los datos de salida mediante X , una variable aleatoria cuyo valor esperado es la cantidad de interés θ . Una simulación independiente (es decir, otra ejecución de la simulación) proporciona una nueva variable aleatoria (digamos Y) independiente de la anterior, con media θ . Este valor θ es llamado parámetro. Este procedimiento continúa hasta obtener, luego de n ejecuciones, un total de n variables aleatorias independientes X_1, \dots, X_n , todas con la misma distribución y media θ . El promedio de estos k valores, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ sirve entonces como estimador, o aproximador, de θ .

¹Un modelo estocástico es una formulación matemática de un fenómeno donde interviene el concepto de aleatoriedad.

²Una simulación del sistema se entiende como un proceso de experimentación virtual.

Consideramos el problema de decidir cuándo detener el estudio de simulación; es decir, el problema de hallar el valor adecuado de n . Para decidir cuándo detenerse, será útil considerar la calidad de nuestro estimador de θ . Además, mostraremos la forma de obtener un intervalo donde podamos afirmar que θ está ahí con cierto grado de confianza. Antes de ver ello revisaremos algunos resultados límite en probabilidad que son pilares de la inferencia estadística.

2.2. EL Teorema central del límite

Primero se estudia lo básico de la teoría de convergencia de variables aleatorias. Sean X y una sucesión $\{X_k : k \geq 1\}$ de v.a. con valores en el espacio euclideo $(\mathbb{R}^d, \|\cdot\|)$, definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se denota con $\|X\| : \Omega \rightarrow \mathbb{R}$ a la función norma de X , definida por $\|X\|(\omega) = \|X(\omega)\|$.

Sea \mathcal{V} el espacio vectorial de las variables aleatorias sobre $(\Omega, \mathcal{F}, \mathbb{P})$ con valores en \mathbb{R}^d . Este espacio \mathcal{V} es de dimensión infinita y, en general, las normas sobre \mathcal{V} no son equivalentes (en el sentido topológico). Por lo tanto, es apropiado distinguir entre las diversas normas resultando distintas nociones de convergencia, lo cual motiva la siguiente definición.

Definición 12. *Decimos que la sucesión $\{X_k, k \in \mathbb{N}\}$ converge a X*

1. *en L_p ($1 \leq p < \infty$) cuando la sucesión $\|X_k - X\|_p = \mathbb{E}(\|X_k - X\|^p)^{1/p}$ converge a 0; la convergencia en L_1 también se llama "la convergencia en media", y la convergencia en L_2 convergencia en media cuadrática";*
2. *casi seguramente (c.s.) cuando $\mathbb{P}\left(\lim_{k \rightarrow \infty} X_k = X\right) = 1$.*
3. *en probabilidad cuando para todo $\epsilon > 0$, la sucesión $\mathbb{P}(\|X_k - X\| > \epsilon)$ converge a 0.*
4. *en ley cuando la ley de X_k converge a la ley de X , es decir, cuando la sucesión $P_{X_k}(E)$ converge a $P_X(E)$ para todo $E \subset \mathbb{R}$ abierto.*

Proposición 5.

1. La convergencia en L_p y la convergencia casi seguramente implican la convergencia en probabilidad.
2. La convergencia en probabilidad implica la convergencia en ley.

Teorema 4.

1. Convergencia monótona: si (V_k) es una sucesión creciente de v.a. reales no negativas que convergen casi seguramente a una v.a. real no negativa V , entonces

$$\lim_{k \rightarrow \infty} \mathbb{E}[V_k] = \mathbb{E} \left[\lim_{k \rightarrow \infty} V_k \right].$$

2. Convergencia dominada (de Lebesgue): Si (X_k) es una sucesión de v.a. que converge casi seguramente a una v.a. X y si existe una v.a. real $Y \in L_1$ dominante: $\|X_k\| \leq Y, k \geq 1$, entonces $X \in L_1$ y la sucesión X_k converge a X en L_1 .

3. Lema de Fatou: Si (X_k) es una sucesión de v.a. reales no negativas, entonces

$$\mathbb{E} \left[\liminf_{k \rightarrow \infty} X_k \right] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[X_k].$$

Proposición 6 (Equivalencias de la convergencia en ley). Las siguientes afirmaciones son equivalentes

1. La sucesión (X_k) de v.a. reales converge en ley a la v.a. real X .
2. La sucesión $\mathbb{E}[f \circ X_k]$ converge a $\mathbb{E}[f \circ X]$ para toda función continua acotada de \mathbb{R} a \mathbb{R} .
3. La sucesión F_{X_k} de funciones de distribución converge puntualmente a la función de distribución de X , F_X , en cada punto de continuidad de F_X .

4. La sucesión φ_{X_k} de funciones características converge puntualmente a la función característica de X , φ_X .

Antes de enunciar el famoso Teorema central del límite se revisan dos teoremas límite conocidos como las Leyes de los grandes números y terminamos con un importante teorema límite para la estadística.

Teorema 5 (Ley débil de los grandes números). Sea (X_k) una sucesión de variables aleatorias reales 2 a 2 no-correlacionadas, cuadrado integrables y con la misma ley. Entonces la sucesión de sumas de Cesaro $\frac{X_1 + \cdots + X_k}{k}$ converge en L_2 a la v.a. real constante $\mathbb{E}[X_1]$.

Teorema 6 (Ley fuerte de los grandes números). Sea (X_k) una sucesión de variables aleatorias reales independientes e integrables, con la misma ley. Entonces la sucesión de sumas de Cesaro $\frac{X_1 + \cdots + X_k}{k}$ converge casi seguramente a la v.a. constante $\mathbb{E}[X_1]$.

Observación 8. También se cumple el recíproco de la ley fuerte de los grandes números: Si (X_k) es una sucesión de v.a. reales independientes y con la misma ley tal que la sumas de Cesaro $\frac{X_1 + \cdots + X_k}{k}$ converge casi seguramente, entonces X_1 es integrable.

Teorema 7 (Teorema central del límite). Sea (X_k) una sucesión de variables aleatorias reales independientes y cuadrado integrables, con la misma ley. Si $\mu = \mathbb{E}[X_1]$ y $\sigma^2 = \text{Var}(X_1)$ entonces la sucesión $\frac{X_1 + \cdots + X_k - k\mu}{\sigma\sqrt{k}}$ converge en ley a una v.a. real normal estándar.

Demostración. Primero se considera el caso $\mu = 0$. Sea $\varphi(t)$ la función característica de X_1 , como X_1 es cuadrado integrable entonces

$$\varphi(t) = 1 + i\mathbb{E}[X_1]t - \frac{\mathbb{E}[X_1^2]}{2}t^2 + \rho(t)t^2 = 1 - \frac{\sigma^2}{2}t^2 + \rho(t)t^2, \quad t \in \mathbb{R},$$

con $\rho(t) \rightarrow 0$ cuando $t \rightarrow 0$. Por independencia, la función característica de $Z_k = \frac{X_1 + \dots + X_k}{\sigma\sqrt{k}}$ es

$$\begin{aligned}\varphi_{Z_k}(t) &= \varphi_{X_1 + \dots + X_k} \left(\frac{1}{\sigma\sqrt{k}} t \right) \\ &= \left[\varphi \left(\frac{1}{\sigma\sqrt{k}} \right) \right]^k \\ &= \left[1 + \frac{1}{k} \left(-\frac{1}{2} t^2 + \rho \left(\frac{t}{\sigma\sqrt{k}} \right) \frac{t^2}{\sigma^2} \right) \right]^k.\end{aligned}$$

Además sabemos que para toda sucesión $(z_k) \subset \mathbb{C}$ convergente se cumple

$$\lim_{k \rightarrow +\infty} \left[1 + \frac{z}{k} \right]^k = \exp \left(\lim_{k \rightarrow \infty} z_k \right),$$

por lo tanto tenemos, para todo $t \in \mathbb{R}$,

$$\begin{aligned}\lim_{k \rightarrow +\infty} \varphi_{Z_k}(t) &= \lim_{k \rightarrow +\infty} \left[1 + \frac{1}{k} \left(-\frac{1}{2} t^2 + \rho \left(\frac{t}{\sigma\sqrt{k}} \right) \frac{t^2}{\sigma^2} \right) \right]^k \\ &= \exp \left(\lim_{k \rightarrow \infty} -\frac{1}{2} t^2 + \rho \left(\frac{t}{\sigma\sqrt{k}} \right) \frac{t^2}{\sigma^2} \right) \\ &= \exp \left(-\frac{1}{2} t^2 \right).\end{aligned}$$

Es decir, el límite de funciones características φ_{Z_k} converge a la función característica de una v.a. real normal estándar Z y así, la sucesión Z_k converge en ley Z , una v.a. real normal estándar. Segundo, si $\mu = \mathbb{E}[X_1]$, tenemos que la sucesión $Y_k = X_k - \mu$ es una sucesión de v.a. independientes cuadrado integrables, con la misma ley y con $\mathbb{E}[Y_1] = 0$ y $\text{Var}(Y_1) = \sigma^2$, es decir se reduce al primer caso. Luego, la sucesión $\frac{Y_1 + \dots + Y_k}{\sigma\sqrt{k}}$, que coincide con la sucesión $\frac{X_1 + \dots + X_k - k\mu}{\sigma\sqrt{k}}$, converge en ley a una v.a. real normal estándar. \square

Teorema 8 (Slutsky). Sean (X_k) , (Y_k) dos sucesiones de v.a. reales. Si Y_k converge en ley a la v.a. real Y y X_k converge en probabilidad a $c \in \mathbb{R}$, entonces se cumple que

1. $X_k + Y_k$ converge en ley a $c + Y$.
2. $X_k \cdot Y_k$ converge en ley a $c \cdot Y$.

3. si además $c \neq 0$, $\frac{Y_k}{X_k}$ converge en ley a $\frac{X}{c}$.

Teorema 9. Sea (X_k) una sucesión de variables aleatorias reales independientes y cuadrado integrables, con la misma ley. Si $\bar{X}_k = \frac{1}{k} \sum_{i=1}^k X_i$ y $S_k = \frac{1}{k-1} \sum_{i=1}^k (X_i - \bar{X}_k)^2$, $k \geq 2$, entonces la sucesión $\frac{\bar{X}_k - \mathbb{E}[X_1]}{S_k/\sqrt{k}}$ converge en ley a una v.a. real normal estándar.

Demostración. Para realizar la prueba combinamos los resultados del Teorema central del límite y el inciso (c) del teorema de Slutsky. Sean las sucesiones

$$Z_k = \frac{\bar{X}_k - \mathbb{E}[X_1]}{\sigma/\sqrt{k}} = \frac{X_1 + \cdots + X_k - k\mathbb{E}[X_1]}{\sigma\sqrt{k}}, \quad Y_k = \frac{S_k}{\sigma}.$$

Sabemos, del Teorema central del límite, que Z_k converge en ley a Z (donde Z es una v.a. normal estándar). Por otro lado, se verifica que Y_k converge en probabilidad a 1. Por lo tanto, del teorema de Slutsky, se tiene

$$\frac{\bar{X}_k - \mathbb{E}[X_1]}{S_k/\sqrt{k}} = \frac{Z_k}{Y_k}$$

converge en ley a $\frac{Z}{1}$. □

2.3. Estimación puntual

Definición 13. Sean X_1, \dots, X_n variables aleatorias, cada una con distribución F (con función de densidad f) y sea θ un parámetro de F (de f). Se dice que $\theta_n = g(X_1, \dots, X_n)$, con g una función continua, es un estimador de θ . Se dice que un estimador θ_n es insesgado si $\mathbb{E}[\theta_n] = \theta$.

Tal como esta enunciado la definición de estimador se pueden tener muchos estimadores para un mismo parámetro, e incluso algunos que no se “aproximen” a nuestro parámetro objetivo, entonces queda claro que lo que buscamos es determinar “buenos estimadores” los cuales se definen en función al parámetro a estimar. Esta es la razón de definir estimador

insesgado.

Definición 14. Sea $\{X_n: n \geq 1\}$ una sucesión de variables aleatorias independientes, cada una con distribución F (con función de densidad f) y $n \geq 1$, se define la media muestral de X_1, \dots, X_n como sigue

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Observación 9. Con las notaciones de la definición precedente se cumple:

1. Sean $\theta = \mathbb{E}[X_1]$ y $\sigma^2 = \text{Var}(X_1)$, entonces θ y σ^2 son parámetros de F (de f).
2. Si $\theta = \mathbb{E}[X_1]$ entonces la media muestral, \bar{X}_n , es un estimador insesgado de θ . En efecto, notemos que la función $g(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$ es continua y que de la definición de \bar{X}_n se sigue

$$\begin{aligned} \mathbb{E}[\bar{X}_n] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \theta \end{aligned} \tag{2.1}$$

Definición 15. Sea $\{X_n: n \geq 1\}$ una sucesión de variables aleatorias independientes, cada una con distribución F (con función de densidad f) y $n \geq 1$, se define la el error cuadrático medio de la media muestral de X_1, \dots, X_n como sigue

$$\text{ECM}(\bar{X}_n) = \text{Var}(\bar{X}_n).$$

Se considera que la distribución F (función de densidad f) determina a una población, así \bar{X}_n es un estimador del parámetro θ al tomar una muestra de dicha población. Así, el

error cuadrático medio nos permite medir la “bondad” de \bar{X}_n como estimador del parámetro θ , interpretado como la media poblacional.

Ahora,

$$\begin{aligned}
 \text{ECM}(\bar{X}_n) &= \mathbb{E}[(\bar{X}_n - \theta)^2], & (\text{donde } \mathbb{E}[\bar{X}_n] = \theta) \\
 &= \text{Var}(\bar{X}_n) \\
 &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i), & (\text{son independientes}) \\
 &= \frac{\sigma^2}{n}, & (\text{pues } \text{Var}(X_i) = \sigma^2) \quad (2.2)
 \end{aligned}$$

Así, la media muestral, \bar{X}_n , es una variable aleatoria con media θ y varianza $\frac{\sigma^2}{n}$. La desigualdad de Chebyshev nos dice que es poco probable que la distancia de una variable aleatoria a su media sea mayor que un múltiplo grande de su desviación estándar, y por lo tanto se tiene que \bar{X}_n es un buen estimador de θ cuando $\frac{\sigma}{\sqrt{n}}$ es pequeño.

La dificultad del uso directo de $\frac{\sigma^2}{n}$ como indicación de lo bien o mal que la media muestral de n datos estima la media poblacional es que por lo general no se conoce la varianza poblacional σ^2 . Así, también necesitamos estimarla. Como

$$\sigma^2 = E[(X_1 - \theta)^2]$$

es el valor esperado del cuadrado de la diferencia entre X_1 y su media (desconocida), al tomar una muestra X_1, \dots, X_n , y \bar{X}_n como estimador de la media poblacional, un estimador natural de σ^2 sería $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, el promedio de los cuadrados de las distancias entre los datos de la muestra y la media estimada. Sin embargo, este estimador no es insesgado, por esta razón tenemos la siguiente definición.

Definición 16. Sea $\{X_n: n \geq 1\}$ una sucesión de variables aleatorias independientes, cada una con distribución F (con función de densidad f) y $n \geq 2$, se define la varianza muestral

de X_1, \dots, X_n como sigue

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Proposición 7. *La varianza muestral es un estimador insesgado, es decir,*

$$\mathbb{E}[S_n^2] = \sigma^2$$

Demostración. Nos servimos de la identidad

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2,$$

para ver que

$$\begin{aligned} (n-1)\mathbb{E}[S_n^2] &= \mathbb{E} \left[(n-1) \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n X_i^2 \right] - n\mathbb{E}[\bar{X}_n^2] \\ &= n\mathbb{E}[X_1^2] - n\mathbb{E}[\bar{X}_n^2] \end{aligned} \tag{2.3}$$

donde la última igualdad se debe a que todas las X_i tienen la misma distribución. Recordando la definición de varianza, para X_1 obtenemos que

$$\begin{aligned} \mathbb{E}[X_1^2] &= \text{Var}(X_1) + (\mathbb{E}[X_1])^2 \\ &= \sigma^2 + \theta^2 \end{aligned}$$

y además para \bar{X}_n , de (2.1) y (2.2), también se tiene que

$$\begin{aligned} \mathbb{E}[\bar{X}_n^2] &= \text{Var}(\bar{X}_n) + (\mathbb{E}[\bar{X}_n])^2 \\ &= \frac{\sigma^2}{n} + \theta^2 \end{aligned}$$

Así, de la ecuación (2.3), obtenemos que

$$(n-1)\mathbb{E}[S_n^2] = n(\sigma^2 + \theta^2) - n \left(\frac{\sigma^2}{n} + \theta^2 \right) = (n-1)\sigma^2,$$

lo cual demuestra el resultado. □

Empleamos la varianza muestral S_n^2 como nuestro estimador de la varianza poblacional σ^2 y a $S_n = \sqrt{S_n^2}$, la llamada desviación estándar muestral, como estimador de σ .

Ahora, en una simulación, estamos interesados en poder estimar el valor de $\theta = \mathbb{E}[X_1]$, con la menor cantidad de datos generados. Por eso, primero debemos elegir un valor aceptable ϵ como cota para la desviación estándar de nuestro estimador, y de este modo debemos continuar generando nuevos datos hasta que con n datos el valor de $\frac{\sigma_n}{\sqrt{n}}$ sea menor que ϵ . Así, planteamos el siguiente procedimiento para determinar el momento de detenernos.

1. Elegir un valor aceptable ϵ para la desviación estándar del estimador.
2. Generar al menos 30 datos.
3. Generar más datos, y detenerse cuando para n valores generados se cumpla que $\frac{S_n}{\sqrt{n}} \leq \epsilon$.
4. La estimación de θ está dada por $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Para utilizar la técnica anterior necesitamos un método para calcular de manera recursiva las sucesivas medias y varianzas muestrales, en vez de calcular todo desde el principio cada vez que se genera un nuevo dato. Las siguientes fórmulas recursivas permiten calcular de manera sucesiva el valor actual de la media y de la varianza muestrales.

Con $S_1^2 = 0$ y $\bar{X}_1 = X_1$ se tiene

$$\bar{X}_{n+1} = \bar{X}_n + \frac{X_{n+1} - \bar{X}_n}{n+1}, \quad (2.4)$$

$$S_{n+1}^2 = \left(1 - \frac{1}{n}\right) S_n^2 + (n+1)(\bar{X}_{n+1} - \bar{X}_n)^2, \quad (2.5)$$

para todo $n \geq 1$.

Com estas consideraciones previas se obtiene el siguiente algoritmo.

Algoritmo 5 (Estimación de la media poblacional).

1. Elegir un valor aceptable ϵ para la desviación estándar del estimador.

2. Generar X_1, X_2, \dots, X_{30} . Sea $n = 30$ y calcular

$$\begin{aligned}\bar{X}_{30} &= \frac{1}{30} \sum_{i=1}^{30} X_i, \\ S_{30}^2 &= \frac{1}{29} \sum_{i=1}^{30} (X_i - \bar{X}_{30})^2.\end{aligned}$$

3. Mientras $\frac{S_n}{\sqrt{n}} > \epsilon$ hacer: Generar X_{n+1} , calcular

$$\begin{aligned}\bar{X}_{n+1} &= \bar{X}_n + \frac{X_{n+1} - \bar{X}_n}{n+1}, \\ S_{n+1}^2 &= \left(1 - \frac{1}{n}\right) S_n^2 + (n+1)(\bar{X}_{n+1} - \bar{X}_n)^2;\end{aligned}$$

y reasignar $n = n + 1$.

4. La estimación de θ está dada por \bar{X}_n .

2.4. Estimación por intervalos de confianza

Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, con media θ y varianza σ^2 . Aunque la media muestral $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ es un estimador eficaz de θ , no esperamos que \bar{X} sea igual a θ , sino que sea “cercano”. Como resultado, a veces es más útil especificar un intervalo para el cual tenemos cierto grado de confianza de que θ esté en él.

Para esto, fijemos el parámetro a estimar θ y consideremos una sucesión, $\{X_n: n \geq 1\}$, de variables aleatorias independientes cada una con función de densidad f . Para cada $A \subset \mathbb{R}$, abierto o cerrado, tenemos

$$\mathbb{P}(X \in A) = \int_A f(x)dx, \quad \mathbb{E}[r(X)] = \int_A r(X)f(x)dx.$$

Definición 17. Un intervalo de confianza al $1 - \alpha$ para el parámetro θ es el intervalo $I_n =]a, b[$ donde $a = a(X_1, \dots, X_n)$ y $b = b(X_1, \dots, X_n)$ son funciones de los datos tal que

$$\mathbb{P}(\theta \in I_n) \geq 1 - \alpha. \quad (2.6)$$

En palabras, $]a, b[$ contiene a θ con probabilidad $1 - \alpha$. Llamamos a $1 - \alpha$ el nivel del intervalo de confianza.

Observación 10. El intervalo I_n es aleatorio mientras que θ es fijo y determinado.

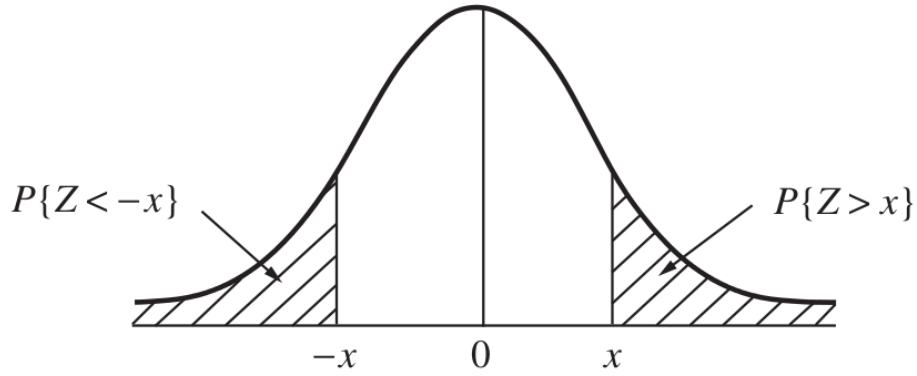


Figura 2.1: Densidad normal estándar

Fuente: Ver figura 8.1 de [6].

Teorema 10 (Intervalo de Confianza Normal). Sean $\{\bar{X}_n: n \geq 1\}$ y $\{S_n: n \geq 1\}$ las sucesiones de medias y desviaciones estándar muestrales asociadas a la sucesión de datos $\{X_n: n \geq 1\}$. Sea Φ la distribución de una variable aleatoria normal estándar y sea $z_{\alpha/2} = \Phi^{-1}(1 - (\alpha/2))$, tal que, $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$ y $\mathbb{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$ donde $Z \sim N(0, 1)$ como en la figura 2.1. Sea

$$I_n = \left[\bar{X}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right], \quad (2.7)$$

entonces

$$\lim_{n \rightarrow \infty} \mathbb{P}(\theta \in I_n) = 1 - \alpha.$$

Demostración. Sea la sucesión $\{Z_n : n \geq 1\}$, definida como sigue

$$Z_n = \frac{\bar{X}_n - \theta}{S_n/\sqrt{n}}.$$

Entonces, del teorema 9, si $Z \sim N(1, 0)$ se tiene que Z_n converge en ley a Z , es decir,

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n < z) = \Phi(z) = \mathbb{P}(Z < z), \quad \text{para todo } z \in \mathbb{R}.$$

Ahora, notemos que $z_{\alpha/2} > 0$ si $0 < \alpha < 1$, y luego se sigue

$$\begin{aligned} \mathbb{P}(\theta \in I_n) &= \mathbb{P}\left(\bar{X}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}} < \theta < \bar{X}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}}\right), \\ &= \mathbb{P}\left(-z_{\alpha/2} < \frac{\theta - \bar{X}_n}{S_n/\sqrt{n}} < z_{\alpha/2}\right), \\ &= \mathbb{P}(-z_{\alpha/2} < Z_n < z_{\alpha/2}), \\ &= \mathbb{P}(Z_n < z_{\alpha/2}) - \mathbb{P}(Z_n < -z_{\alpha/2}), \end{aligned}$$

de donde tomando límite, cuando $n \rightarrow \infty$, se obtiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\theta \in I_n) &= \lim_{n \rightarrow \infty} \mathbb{P}(Z_n < z_{\alpha/2}) - \mathbb{P}(Z_n < -z_{\alpha/2}), \\ &= \mathbb{P}(Z < z_{\alpha/2}) - \mathbb{P}(Z \leq -z_{\alpha/2}), \\ &= \mathbb{P}(Z < z_{\alpha/2}) - \mathbb{P}(Z \geq z_{\alpha/2}), \\ &= \mathbb{P}(Z < z_{\alpha/2}) - (1 - \mathbb{P}(Z < z_{\alpha/2})), \\ &= 2 \cdot \Phi(z_{\alpha/2}) - 1 \\ &= 2 \cdot \Phi(\Phi^{-1}(1 - (\alpha/2))) - 1 \\ &= 1 - \alpha. \end{aligned}$$

Además, también se ha probado que

$$\mathbb{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = \mathbb{P}(Z < z_{\alpha/2}) - \mathbb{P}(Z \leq -z_{\alpha/2}) = 1 - \alpha.$$

□

Observación 11. Para aclarar el significado de un “intervalo de confianza al $1 - \alpha$ (del $100(1 - \alpha)$ por ciento)”, consideremos, por ejemplo, el caso en que $\alpha = 0,05$, de modo que

$z_{\alpha/2} = 1,96$. Ahora, antes de observar los datos, será cierto, con probabilidad (aproximadamente) igual a 0,95, que la media muestral \bar{X}_n y la desviación estándar muestral S_n serán tales que θ estará entre $\bar{X}_n - 1,96 \frac{S_n}{\sqrt{n}}$ y $\bar{X}_n + 1,96 \frac{S_n}{\sqrt{n}}$. Después de observar que \bar{X}_n y S_n sean iguales a \bar{x}_n y s_n , respectivamente, ya no hablamos de la probabilidad de que θ esté en el intervalo $\bar{x}_n - 1,96 \frac{s_n}{\sqrt{n}}$ y $\bar{x}_n + 1,96 \frac{s_n}{\sqrt{n}}$, pues ahora está o no está ahí. Sin embargo, estamos “95 por ciento seguros” de que en esta situación está dentro de dicho intervalo (pues sabemos que, a largo plazo, tales intervalos contendrán la media el 95 por ciento de las veces).

Ahora, en una simulación, dado α estamos interesados en poder estimar un intervalo de confianza del $100(1 - \alpha)$ por ciento para θ , con la menor cantidad de datos generados. Por ello, primero debemos elegir un valor aceptable ϵ como cota para la longitud de nuestro intervalo, que a la vez es mayor que el error de aproximación, y de este modo debemos continuar generando nuevos datos hasta que con n datos la longitud del intervalo de confianza $2z_{\alpha/2} \frac{S_n}{\sqrt{n}}$ sea menor que ϵ . Así, planteamos el siguiente algoritmo.

Algoritmo 6 (Estimación por intervalo de confianza).

1. Elegir un valor aceptable ϵ para el error de aproximación del estimador.
2. Generar X_1, X_2, \dots, X_{30} . Sea $n = 30$ y calcular

$$\bar{X}_{30} = \frac{1}{30} \sum_{i=1}^{30} X_i,$$

$$S_{30}^2 = \frac{1}{29} \sum_{i=1}^{30} (X_i - \bar{X}_{30})^2.$$

3. Mientras $2z_{\alpha/2} \frac{S_n}{\sqrt{n}} > \epsilon$ hacer: Generar X_{n+1} , calcular

$$\bar{X}_{n+1} = \bar{X}_n + \frac{X_{n+1} - \bar{X}_n}{n+1},$$

$$S_{n+1}^2 = \left(1 - \frac{1}{n}\right) S_n^2 + (n+1)(\bar{X}_{n+1} - \bar{X}_n)^2;$$

y reasignar $n = n + 1$.

4. El intervalo de confianza del $100(1 - \alpha)\%$, con longitud menor que ϵ , es

$$\left[\bar{X}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right].$$

2.5. Método por Monte Carlo

Un problema clásico del cálculo infinitesimal es el cálculo de integrales, es decir, dada una función $g: [a, b] \rightarrow \mathbb{R}$ hallar

$$\int_a^b g(x) dx.$$

El análisis matemático permite garantizar la existencia de dicha integral, más específicamente: Sea $g: [a, b] \rightarrow \mathbb{R}$ una función acotada, si g es continua en $[a, b]$ entonces g es integrable en $[a, b]$.

Supongamos que θ es la integral a calcular, es decir,

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

Sea U una variable aleatoria distribuida uniformemente sobre $]a, b[$, entonces podemos expresar la integral de la siguiente forma

$$\theta = \int_a^b (b - a)g(x) \cdot \frac{dx}{b - a} = \mathbb{E}[(b - a)g(U)]$$

Por lo tanto, del capítulo anterior, podemos aproximar θ generando una sucesión de variables aleatorias $g(U_1), g(U_2), \dots, g(U_n)$ independientes e idénticamente distribuidas (esto es generando números aleatorios U_i) y considerar como estimador de θ a la variable aleatoria dada por la siguiente ecuación

$$\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^n (b - a)g(X_i).$$

Además, considerando S_n^2 como la varianza muestral dada por la ecuación

$$S_n^2 = \frac{1}{n - 1} \sum_{i=1}^n ((b - a)g(X_i) - \bar{\theta}_n)^2$$

se obtiene que el intervalo de confianza al $1 - \alpha$, del $100(1 - \alpha) \%$, para θ es

$$\left[\bar{\theta}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{\theta}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right].$$

Así tenemos los siguientes algoritmos:

Algoritmo 7 (El método de integración por Monte Carlo).

1. Elegir n grande.
2. Generar X_1, \dots, X_n valores según una distribución uniforme sobre $]a, b[$.
3. Calcular

$$\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^n (b-a)g(X_i)$$

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n ((b-a)g(X_i) - \theta)^2}$$

También podemos diseñar un algoritmo, usando una fórmulas recursivas análogas a las establecidas en las ecuaciones (2.4) y (2.5), de modo que solo se haga los cálculos necesarios para obtener un intervalo de confianza pequeño como se detalla a continuación.

Algoritmo 8 (El método de integración por Monte Carlo con intervalo de confianza).

1. Elegir un valor aceptable ϵ para el error de aproximación de la integral.
2. Generar X_1, X_2, \dots, X_{30} valores de una variable aleatoria uniforme sobre $]a, b[$. Sea $n = 30$ y calcular

$$\theta_{30} = \frac{1}{30} \sum_{i=1}^{30} (b-a)g(X_i),$$

$$S_{30}^2 = \frac{1}{29} \sum_{i=1}^{30} ((b-a)g(X_i) - \theta_{30})^2.$$

3. Mientras $2z_{\alpha/2} \frac{S_n}{\sqrt{n}} > \epsilon$ hacer: Generar X_{n+1} un valor de una variable aleatoria uniforme sobre $]a, b[$, calcular

$$\theta_{n+1} = \theta_n + \frac{(b-a)g(X_{n+1}) - \theta_n}{n+1},$$

$$S_{n+1}^2 = \left(1 - \frac{1}{n+1}\right) S_n^2 + \frac{1}{n+1} (\theta_{n+1} - \theta_n)^2;$$

y reasignar $n = n + 1$.

4. El intervalo de confianza del $100(1 - \alpha) \%$, con longitud menor que ϵ , es

$$\left[\bar{\theta}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{\theta}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right].$$

La aproximación de la integral de g sobre $[a, b]$ es

$$\bar{\theta}_n.$$

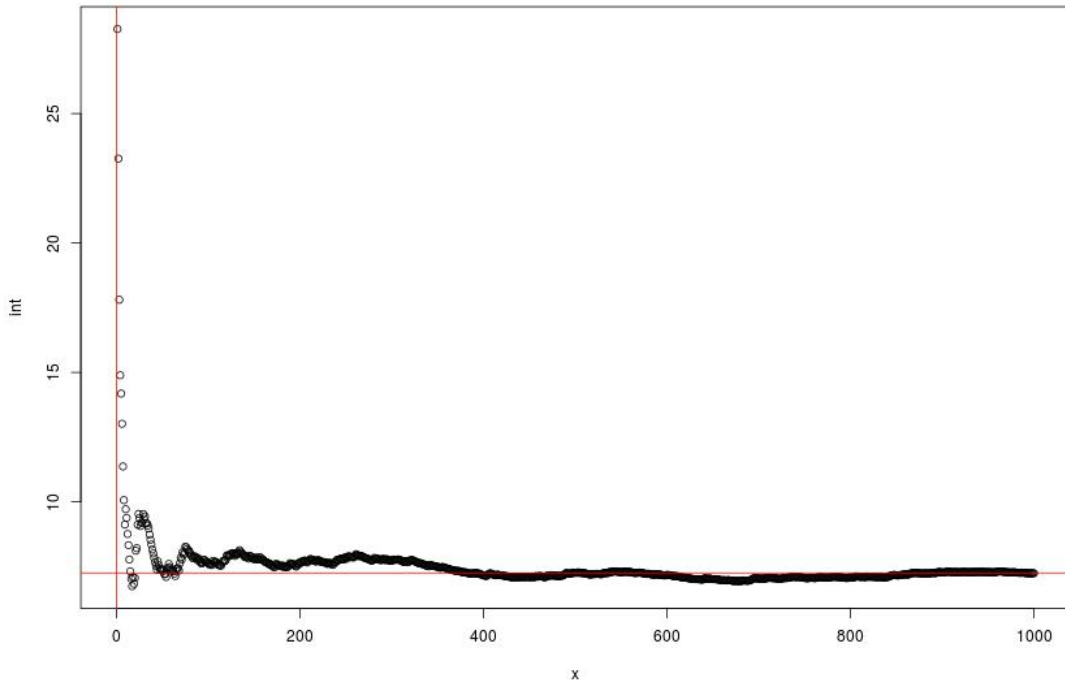


Figura 2.2: Integral de $\exp(x)$ para $n = 10^4$.

Ejemplo 6. Sea $f(x) = e^x$, $x \in]-2, 2[$, queremos hallar

$$\theta = \int_{-2}^2 e^x dx$$

Primero, mediante cálculo directo obtenemos

$$\theta = e^x \Big|_{-2}^2 = e^2 - e^{-2} = 7,25372081569404$$

Segundo, mediante el algoritmo 8 (ver programa 2), para $n = 2168424$ iteraciones se obtiene que la aproximación de la integral es

$$\bar{\theta}_n = 7,25499352101258$$

y el respectivo intervalo de confianza de nivel 95 % es

$$I_n = [7,19619523149124 - 7,31379181053392].$$

El error de esta aproximación es 0,01754555 %.

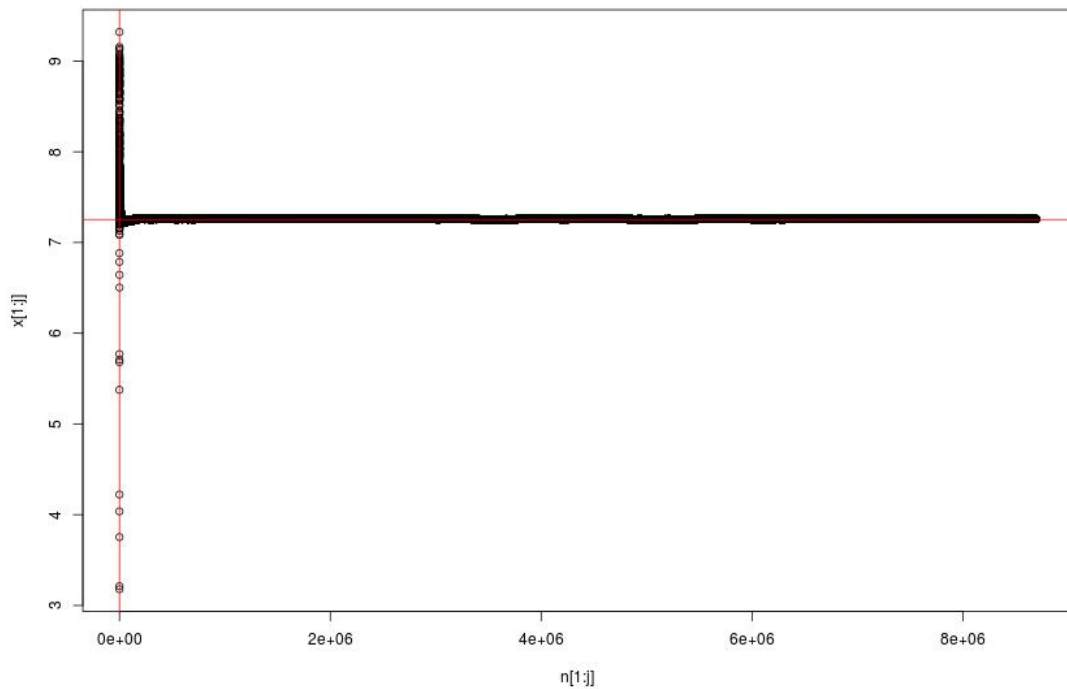


Figura 2.3: Integral de $\exp(x)$.

La “forma de convergencia” del algoritmo de integración por Monte Carlo se puede observar en la gráfica 2.2 donde la línea horizontal roja marca el valor teórico de la integral.

Ejemplo 7 (Una integral no elemental). Sea $f(x) = e^{-x^2/2}$, $x \in]-3, 3[$, queremos hallar

$$\theta = \int_{-3}^3 e^{-x^2/2} dx.$$

Esta integral es conocida y se calcula mediante tabla, por tanto podemos tomar como “valor teórico”

$$\theta = 2,49910838980711$$

El método de integración por Monte Carlo (ver programa 3), para $n = 167964$ iteraciones, nos da que la aproximación es

$$\bar{\theta}_n = 2,49413002212611$$

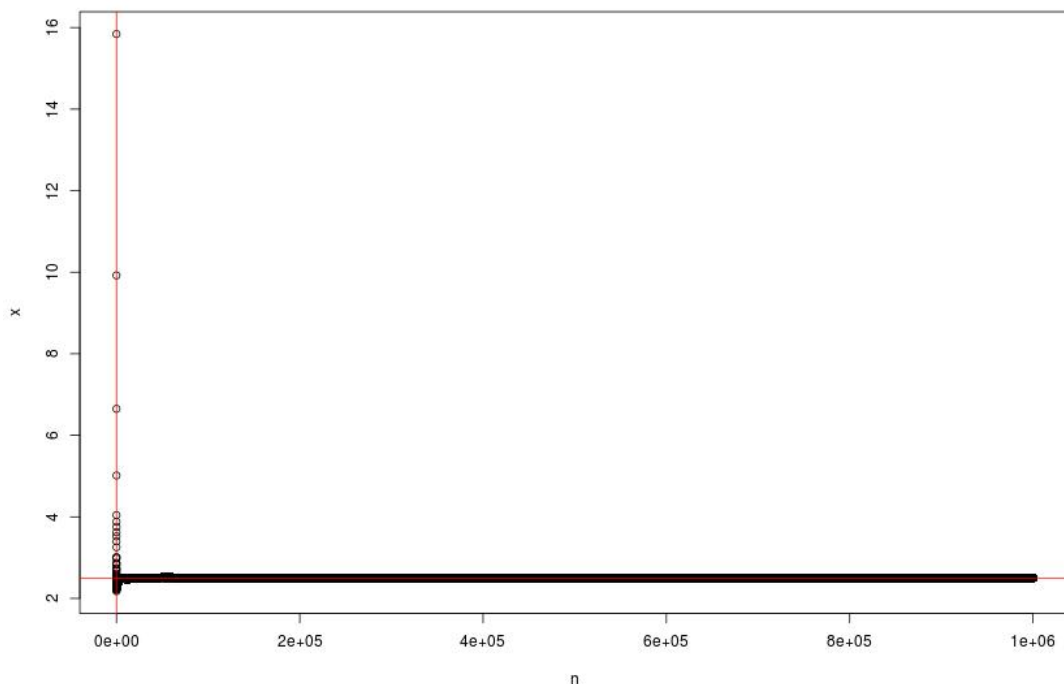


Figura 2.4: Integral de $\exp(-x^2/2)$.

y el respectivo intervalo de confianza de nivel 95 % es

$$I_n = [2,4745306785623 - 2,51372936568993].$$

El error de esta aproximación es 0,1992058 %.

En la figura 2.4 se observa (la línea horizontal roja indica el valor teórico) el proceso de cálculo del método de integración por Monte Carlo para este caso.

Finalmente para ver diversas aplicaciones y variantes de este método puede revisarse el capítulo 6 del libro [1]. También hay otros libros que exploran métodos más sofisticados que el descrito, por ejemplo en el libro [6] se desarrolla el, muy celebrado y usado, método de Monte Carlo con cadenas de Markov.

Capítulo 3

Método de Monte Carlo con Cadenas de Markov

En general, es difícil simular el valor de un vector aleatorio X cuyas variables aleatorias componentes son dependientes. En este capítulo presentamos un método poderoso para generar un vector cuya distribución es aproximadamente la de X . Este método, llamado método de Monte Carlo con cadenas de Markov, tiene la ventaja adicional de que la función de masa (o densidad) de X puede estar dada salvo una constante multiplicativa, lo cual es de gran importancia en las aplicaciones, como veremos más adelante.

En la sección 3.1 presentamos los resultados necesarios sobre las cadenas de Markov. En la sección 3.2 explicamos el algoritmo de Hastings-Metropolis para construir una cadena de Markov con una función de masa de probabilidad dada como distribución límite. Además, vemos un caso particular de este algoritmo, el muestreador de Gibbs, que es tal vez el método de Monte Carlo con cadenas de Markov más utilizado. En la sección 3.3 vemos una aplicación del método de Monte Carlo con cadenas de Markov al modelo de Ising.

3.1. Cadenas de Markov

Consideremos una colección de variables aleatorias X_0, X_1, \dots donde X_n es estado del sistema en el instante n y supona que el conjunto de los posibles valores de las X_n son el conjunto $1, \dots, N$. Si hay un conjunto de puntos P_{ij} , $i, j = 1, \dots, N$ tales que siempre el proceso esta en estado i , entonces, de manera independiente a los estados pasados, la probabilidad de que el siguiente estado sea j es P_{ij} , entonces decimos que la colección $X_n, n \geq 0$ es una **cadena de markov** con probabilidades de transición $P_{ij}, i, j = 1 \dots, N$. Como el proceso debe estar en algún estado después de dejar el estado i , estas probabilidades de transición satisfacen

$$\sum_{j=1}^N P_{ij} = 1, \quad i = 1, \dots, N$$

Una cadena de Markov es irreducible si para cada par de estados i y j existe una probabilidad positiva de que, partiendo del estado i , el proceso llega a estar en el estado j . Para una cadena de Markov irreducible, sea π_j la fracción de tiempo, a largo plazo, que el proceso pasa en el estado j (se puede mostrar que π_j existe y es constante, con probabilidad 1, de manera independiente al estado inicial). Es posible mostrar que las cantidades $\pi_j, j = 1, \dots, N$, son la única solución del siguiente conjunto de ecuaciones lineales:

$$\pi_j = \sum_{i=1}^N \pi_i P_{ij}, \quad j = 1, \dots, N \quad \text{y} \quad \sum_{j=1}^N \pi_j = 1 \quad (3.1)$$

Observación 12. Las ecuaciones 3.1 tiene una interpretación heurística: π_i es la fracción de tiempo que la cadena de Markov está en el estado i y como cada transición que sale del estado i está en el estado j con probabilidad P_{ij} , se tiene que $\pi_i P_{ij}$ es la fracción de tiempo en la cual la cadena de Markov acaba de entrar al estado j desde el estado i . Por lo tanto, la primera ecuación establece el hecho intuitivamente claro de que la fracción de tiempo en la cual la cadena de Markov acaba de entrar al estado j es igual a la suma, sobre todos los estados i , de la fracción de tiempo en que acaba de entrar al estado j proveniente del estado i . Por supuesto, la segunda ecuación de dice que la suma de la fracción de tiempo en la cual la cadena está en el estado j , sobre todo j , debe ser igual a 1.

Con frecuencia π_j se llaman probabilidades estacionarias de la cadena de Markov, ya

que si el estado inicial de la cadena de Markov se distribuye de acuerdo con π_j , entonces $P\{X_n = j\} = \pi_j$ para todo instante n y estado j .

Una propiedad importante de las cadenas de Markov es que para cualquier función h sobre el espacio de estados, con probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N \pi_j h(j) \quad (3.2)$$

Lo anterior se debe a que si $p_j(n)$ es la fracción de tiempo que la cadena está en el estado j entre los instantes $1, \dots, n$, entonces

$$\frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N h(j) p_j(n) \rightarrow \sum_{j=1}^N h(j) \pi_j$$

con frecuencia, la cantidad π_j se puede interpretar como la probabilidad límite de que la cadena esté en el estado j . Para precisar las condiciones en las que es válida esta interpretación, primero necesitamos la definición de una cadena de Markov aperiódica.

Definición 18. Una cadena de Markov irreducible es aperiódica si para algún $n \geq 0$ y algún estado j ,

$$P\{X_n = j | X_0 = j\} > 0 \quad y \quad P\{X_{n+1} = j | X_0 = j\} > 0$$

Se puede mostrar que si la cadena de Markov es irreducible y aperiódica, entonces

$$\pi_j = \lim_{n \rightarrow \infty} P\{X_n = j\} \quad j = 1, \dots, N$$

A veces hay una forma más fácil de determinar las probabilidades estacionarias que resolver el conjunto de ecuaciones 3.1. Suponga que existen números positivos $x, j = 1, \dots, N$ tales que

$$x_i P_{ij} = x_j P_{ji} \quad \text{para } i \neq j, \quad \sum_{j=1}^N x(j) = 1$$

Al sumar las ecuaciones anteriores sobre todos los estados i se obtiene

$$\sum_{j=1}^N x_i P_{ij} = x_j \sum_{j=1}^N P_{ji} = x_j$$

y como $\{\pi_j, j = 1, \dots, N\}$ son la única solución de 3.1, esto implica que

$$\pi_j = x_j$$

Cuando $\pi_i P_{ij} = \pi_j P_{ji}$ para $i \neq j$, la cadena de Markov es **reversible en el tiempo**, pues se puede mostrar que bajo esta condición, si el estado inicial se elige de acuerdo con las probabilidades π_j y se inicia en cualquier momento la sucesión de estados que va hacia atrás en el tiempo, ésta será también una cadena de Markov con probabilidades de transición P_{ij} .

Sumpongamos ahora que queremos generar el valor de una variable aleatoria X con función de masa de probabilidad $P\{X = j\} = p_j, j = 1, \dots, N$ entonces podríamos generar de manera aproximada tal variable aleatoria ejecutando la cadena durante n pasos para obtener el valor de X_n donde n es grande. Además, si nuestro objetivo era generar muchas variables aleatorias distribuidas de acuerdo con $p_j, j = 1, \dots, N$, para poder estimar $E[h(X)] = \sum_{j=1}^N h(j)p_j$, entonces también podríamos estimar esta cantidad mediante el estimador $\frac{1}{n} \sum_{i=1}^n h(X_i)$. Sin embargo, como los primeros estados de la cadena de Markov pueden tener una fuerte influencia del estado inicial elegido, en la práctica es común desechar los primeros k estados, para algún valor adecuado de k . Es decir, se utiliza el estimador $\frac{1}{n-k} \sum_{i=k+1}^n h(X_i)$. Es difícil saber con exactitud qué tan grande debe ser el valor de k [aunque el lector avanzado puede consultar Aarts y Korst (1989) para obtener algunos resultados útiles al respecto] y por lo común sólo se utiliza la intuición (que en general funciona, pues la convergencia queda garantizada sin importar el valor utilizado).

Una cuestión importante es cómo utilizar la cadena de Markov simulada para estimar el error cuadrático medio del estimador. Es decir, si

$$\hat{\theta} = \frac{1}{n-k} \sum_{i=k+1}^n h(X_i),$$

cómo estimar

$$\text{ECM}(\hat{\theta}, \theta) = E \left[\left(\hat{\theta} - \sum_{j=1}^N h(j)p_j \right)^2 \right].$$

Una forma es el método de **medias por lotes**, que funciona como sigue. Se separan los $n-k$ estados generados en s lotes de tamaño r , donde $s = \frac{n-k}{r}$ es entero, y sea $Y_n, j = 1, \dots, s$

el promedio del j -ésimo lote. Es decir,

$$Y_j = \frac{1}{r} \sum_{i=k+(j-1)r+1}^{k+jr} h(X_i), \quad j = 1, \dots, s$$

Ahora, consideramos las $Y_j, j = 1, \dots, s$ como si fuesen independientes e idénticamente distribuidas con varianza σ^2 y tomamos su varianza muestral

$$\sum_{j=1}^s \frac{(Y_j - \bar{Y})^2}{s-1}$$

como estimador de σ^2 . La estimación del error cuadrático medio es

$$\frac{\widehat{\sigma}^2}{s}.$$

El valor adecuado de r depende de la cadena de Markov simulada. Mientras más cerca estén las $X_i, i \geq 1$, de ser independientes e idénticamente distribuidas, menor será el valor de r .

En las dos secciones siguientes mostraremos, para un conjunto dado de números positivos $b_j, j = 1, \dots, N$, la forma de construir una cadena de Markov cuyas probabilidades limite sean

$$\pi_j = \frac{b_j}{\sum_{i=1}^N b_i}$$

.

3.2. El algoritmo de Hastings-Metropolis

Sean $b(j), j = 1, \dots, m$ números positivos, y sea $B = \sum_{j=1}^m b(j)$. Suponga que m es grande, que B es difícil de calcular y que queremos simular una variable aleatoria(o una sucesión de variables aleatorias) con función de masa de probabilidad

$$\pi(j) = \frac{b(j)}{B}, \quad j = 1, \dots, m$$

Una forma de simular una sucesión de variables aleatorias cuyas distribuciones convergen a $\pi(j), j = 1, \dots, m$, consiste en determinar una cadena de Markov que sea fácil de simular

y cuyas probabilidades límites sean las π_j . El **algoritmo de Hastings-Metropolis** proporciona un método para realizar esta tarea. Construye una cadena de Markov reversible en el tiempo con las deseadas probabilidades límites, de la manera siguiente.

Sea Q una matriz de probabilidades de transición de Markov irreducible sobre los enteros $j = 1, \dots, m$, donde $q(i, j)$ representa el elemento de Q en la fila i y la columna j . Definimos una cadena de Markov $\{X_n, n \geq 0\}$ como sigue. Cuando $X_n = i$, se genera una variable aleatoria X tal que $P\{X = j\} = q(i, j), j = 1, \dots, m$. Si $X = j$, entonces X_{n+1} es igual a j con probabilidad $\alpha(i, j)$ y es igual a i con probabilidad $1 - \alpha(i, j)$. En estas condiciones, es fácil ver que la sucesión de estados formará una cadena de Markov con probabilidades de transición P_{ij} dadas por

$$P_{i,j} = q(i, j)\alpha(i, j), \quad \text{si } i \neq j$$

$$P_{i,i} = q(i, i) + \sum_{k \neq i} q(i, k)(1 - \alpha(i, k))$$

Esta cadena de Markov será reversible en el tiempo y tendrá probabilidades estacionarias $\pi(j)$ si

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \text{ para } i \neq j$$

que es equivalente a

$$\pi(i)q(i, j)\alpha(i, j) = \pi(j)q(j, i)\alpha(j, i)$$

Ahora es fácil verificar que esto se satisface si consideramos

$$\alpha(i, j) = \min \left(\frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}, 1 \right) = \min \left(\frac{b(j)q(j, i)}{b(i)q(i, j)}, 1 \right)$$

Para verificar, observe que si $\alpha(i, j) = \frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}$ entonces $\alpha(j, i) = 1$ y viceversa.

Observación 13. *El valor de B no es necesario para definir la cadena de Markov, pues bastan los valores $b(j)$. Además, casi siempre ocurre que $\pi(j), j = 1, \dots, m$ no sólo serán probabilidades estacionarias sino también probabilidades límites (de hecho, una condición suficiente es que $P_{i,i} > 0$ para alguna i).*

El algoritmo de Hastings-Metropolis para generar una cadena de Markov reversible en el tiempo, cuyas probabilidades límites son $\pi(j) = \frac{b(j)}{B}$, $j = 1, \dots, m$ puede escribirse en forma algorítmica de la siguiente manera

Algoritmo 9 (El algoritmo de Hastings-Metropolis).

1. Elegir una matriz Q de probabilidades de transición, de Markov, irreducible, con probabilidades de transición $q(i, j)$, $i, j = 1, \dots, m$. Además, elegir algún valor entero k entre 1 y m .
2. Elegir $n = 0$ y $X_0 = k$.
3. Generar una variable aleatoria X tal que $P\{X = j\} = q(X_n, j)$ y generar un número aleatorio U .
4. Si $U < \frac{b(X)q(X, X_n)}{b(X_n)q(X_n, X)}$ entonces $NS = X$; en caso contrario, $NS = X_n$.
5. hacer $n = n + 1$ y $X_n = NS$.
6. Ir al paso 3.

3.2.1. El muestreador de Gibbs

La versión más utilizada del algoritmo de Hastings-Metropolis es el muestreador de Gibbs. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con función de masa de probabilidad (función de densidad de probabilidad en el caso continuo) $p(x)$, la cual sólo está determinada salvo una constante multiplicativa, y suponga que queremos generar un vector aleatorio cuya distribución sea la de la distribución condicional de X , dado que $X \in A$ para algún conjunto A , Es decir, queremos generar un vector aleatorio con función de masa

$$f(x) = \frac{p(x)}{P\{X \in A\}}, \text{ para } x \in A$$

El muestreador de Gibbs supone que para cualquier $i, i = 1, \dots, n$ y cualesquiera valores $x_j, j \neq i$, podemos generar una variable aleatoria X con la función de masa de probabilidad

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$$

Esto opera considerando una cadena de Markov con estados $x = (x_1, \dots, x_i, \dots, x_n) \in A$; luego se utiliza el algoritmo de Hastings-Metropolis con las probabilidades de transición de Markov definidas como sigue. Siempre que el estado actual sea x , se genera una coordenada que tiene la misma probabilidad de ser cualquiera de los valores $1, \dots, n$. Si la coordenada i es la elegida, entonces se genera una variable aleatoria X con función de masa de probabilidad $P\{X_i = x | X_j = x_j, j \neq i\}$ y si $X = x$, entonces se considera el estado $y = (x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$ para la transición. En otras palabras, el muestreador de Gibbs emplea el algoritmo de Hastings-Metropolis con

$$q(x, y) = \frac{1}{n} P\{X_i = x | X_j = x_j, j \neq i\} = \frac{1}{n} \frac{p(y)}{P\{X_j = x_j, j \neq i\}}$$

Como la función de masa objetivo es f , el vector y se acepta como el nuevo estado con probabilidad

$$\alpha(i, j) = \min\left(\frac{f(y)q(y, x)}{f(x)q(x, y)}, 1\right)$$

Como para $x \in A$ y $y \in A$, se tiene que

$$\frac{f(y)q(y, x)}{f(x)q(x, y)} = \frac{f(y)p(x)}{f(x)p(y)} = 1$$

mientras que para $x \in A$ y $y \notin A$

$$\frac{f(y)q(y, x)}{f(x)q(x, y)} = 0$$

vemos que el siguiente estado es y si $y \in A$ o sigue siendo x si $y \notin A$.

En resumen, vemos que la cadena de Markov reversible en el tiempo, con probabilidades estacionarias dadas por f , generada por el muestreador de Gibbses la siguiente.

Algoritmo 10 (El muestreador de Gibbs).

1. Sea $x = (x_1, \dots, x_n)$ un vector en A para el cual $p(x) > 0$.
2. Sea I una variable aleatoria que toma uno de los valores $1, \dots, n$, donde cada valor tiene la misma probabilidad de ocurrir.
3. Si $I = i$ generar el valor de una variable aleatoria X tal que

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\}$$

.

4. Si $X = x$ y $(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \in A$, entonces el nuevo valor de x_i es igual a x . En caso contrario, se mantiene el valor de x_i .

5. Regresar al paso 2..

Así, en cada paso se elige una de las variables X_i al azar y se genera una variable aleatoria con la distribución condicional de X_i dado que $X_j = x_j, j \neq i$. Si el nuevo vector, donde este nuevo valor reemplaza a x_i , está en A , entonces es el siguiente estado de la cadena; si el vector no está en A , entonces el estado permanece sin cambios.

3.3. El modelo de Ising

El modelo de Ising es una distribución de probabilidad en $\mathcal{X} = \{-1, +1\}^V$ donde V es el conjunto de vértices de un grafo $G = (V; E)$. El valor $\sigma(v)$ se llama spin en v . La interpretación física es que los imanes, cada uno con una de las dos posibles orientaciones representadas por $+1$ y -1 , se colocan en los vértices del gráfico; una configuración específica a las orientaciones de estos imanes.

El modelo de Ising de vecino más cercano es el sistema de spin más estudiado. En este sistema, la energía de una configuración se denota como

$$H(\sigma) = - \sum_{\substack{v, v' \in V \\ v \sim v'}} \sigma(v)\sigma(v'),$$

donde $\sigma = (\sigma(v))_{v \in V} \in \mathcal{X}$. La energía aumenta con el número de pares de vecinos cuyos spins no coinciden.

La **distribución de Gibbs** correspondiente a la energía H es la distribución de probabilidad sobre \mathcal{X} denotada por

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$$

Aquí la **función de partición** $Z(\beta)$ es la constante de normalización necesaria para hacer una distribución de probabilidad:

$$Z(\beta) := \sum_{\sigma \in \mathcal{X}} e^{-\beta H(\sigma)}$$

La interpretación física nos dice que β es la inversa de la temperatura.

3.3.1. Transición de fase

El modelo descrito anteriormente más precisamente se conoce como modelo de Ising con condiciones de frontera libre, pero cambiando el conjunto de índices sobre el cual se computa la energía se pueden considerar condiciones de frontera periódica o una condición de frontera predeterminada (Como en la figura 3.1).

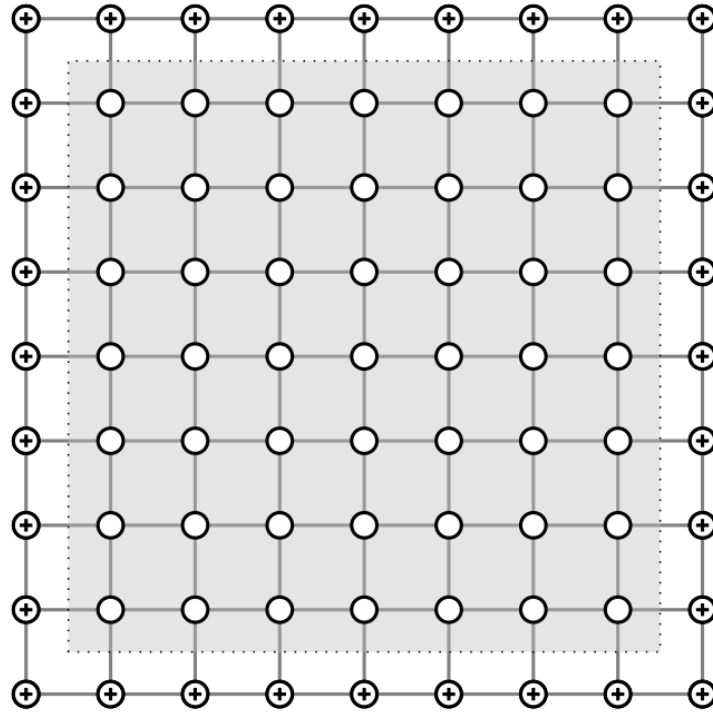


Figura 3.1: El modelo Ising en el un grafo (caja sombreada) con condiciones de frontera +.

El parámetro $\beta > 0$ determina la influencia de la función energética. A una temperatura infinita ($\beta = 0$), la función de energía H no juega ningún papel y μ es la distribución uniforme en \mathcal{X} . En este caso, no hay interacción entre los spins en diferentes vértices y las variables aleatorias $\{\sigma(v)\}_{v \in V}$ son independientes. A medida que aumenta $\beta > 0$, también aumenta el sesgo de μ hacia configuraciones de baja energía. De hecho, esto se puede escribir como lo indica el siguiente resultado.

Teorema 11 (Transición de fase del modelo de Ising en \mathbb{Z}^d con $d \geq 2$, [3]). *Existe un beta crítico $0 < \beta_c < +\infty$ de tal manera que*

- Si $\beta < \beta_c$

$$|\mathcal{G}_\beta| = 1$$

- Si $\beta > \beta_c$

$$|\mathcal{G}_\beta| > 1$$

donde \mathcal{G}_β es el conjunto de medidas de Gibbs en \mathbb{Z}^d formado por las medidas que son puntos límite del conjunto de las medidas de Gibbs en las cajas finitas que tienden a cubrir \mathbb{Z}^2 considerando diferentes condiciones de contorno sobre estas cajas.

En el presente trabajo no se tiene por objeto precisar como se construyen estas medidas de Gibbs en \mathbb{Z}^d pero si se puede comentar que hay la conjetura ([7]) que el beta crítico en dimensión $d = 2$ es

$$\beta_c = \frac{\log(1 + \sqrt{2})}{2},$$

y por lo tanto se tendría una temperatura crítica $T_c = 1/\beta_c$.

Es por ello que esta con un algoritmo de simulación de medidas de Gibbs en una caja regularmente grande puede dar indicios de esta transición de fase cuando para temperaturas bajas ($\beta > \beta_c$) se observa la formación de clusters (conglomerados) de spins -1 (o $+1$) mientras que para temperaturas altas $\beta < \beta_c$ tiende a observarse una configuración de spins más parecida a un sorteo independiente uniforme de -1 y $+1$ (Figura 3.2).

3.3.2. Simulación de la medida de Gibbs

La dinámica de Glauber para la distribución de Gibbs μ se mueve desde una configuración inicial σ seleccionando un vértice v uniformemente al azar de V y luego generando una nueva configuración de acuerdo con μ condicionada al conjunto de configuraciones que coinciden con σ en vértices diferentes de v .

Se puede comprobar que la μ -probabilidad condicional de tener el spin $+1$ en v ,

$$p(\sigma, v) := \mu(\sigma'(v) = +1 | \sigma'(v') = \sigma(v'), \forall v' \sim v),$$

está dada por

$$p(\sigma, v) = \frac{e^{\beta S(\sigma, v)}}{e^{\beta S(\sigma, v)} + e^{-\beta S(\sigma, v)}} = \frac{1}{1 + e^{-2\beta S(\sigma, v)}}$$

donde $S(\sigma, v) := \sum_{v': v' \sim v} \sigma(v')$. Hay que tener en cuenta que $p(\sigma, v)$ depende solo de los spins en los vértices adyacentes a v . Por lo tanto, se tiene una matriz de probabilidad de transición en \mathcal{X} definida por

$$P(\sigma, \sigma') = \frac{1}{|V|} \sum_{v \in V} \frac{e^{\beta \sigma'(v) S(\sigma, v)}}{e^{\beta \sigma'(v) S(\sigma, v)} + e^{-\beta \sigma'(v) S(\sigma, v)}} \cdot 1_{\{\sigma'(v) = \sigma(v), v' \neq v\}}$$

Teorema 12 ([5]). *La cadena de Markov asociada a la matriz de probabilidad de transición P admite una distribución estacionaria y esta es la distribución de Gibbs μ .*

Por lo tanto se tiene el siguiente algoritmo para generar configuraciones asociadas a la medida de Gibbs.

Algoritmo 11 (Algoritmo de simulación de la medida de Gibbs).

1. Sea $\sigma_0 \in \mathcal{X}$ de tal manera que acuerdo con $(\sigma_0(v))_{v \in V}$ sean variables i.i.d. $\sim \text{Unif}(\{-1, 1\})$.

Hacer $X_0 = \sigma_0$.

2. Elegir $\sigma_1 \in \mathcal{X}$ según la distribución $P(\sigma_0, \cdot)$:

Fijar $v \sim \text{Unif}(V)$.

Simular $u \sim \text{Unif}([0, 1])$

■ Si $u < p(\sigma_0, v)$ hacer $\sigma_o(v) = +1$

■ Caso contrario $\sigma_o(v) = -1$

Hacer $\sigma_1 = \sigma_0$ y $X_1 = \sigma_1$.

3. Regresar al paso 2.

En la Figura 3.2 se simula una configuración según la medida de Gibbs sobre un grafo cuyo conjunto de vértices es

$$V = \{1, 2, \dots, 200\}^2$$

y cuyo conjunto de aristas es

$$E = \{\{v, v'\} \in V^2 : |v_1 - v'_1| + |v_2 - v'_2| = 1\}.$$

En esa figura se puede apreciar el cambio en las configuraciones típicas al cambiar la temperatura (o equivalentemente el β) como fue descrito al hablar de transición de fase.

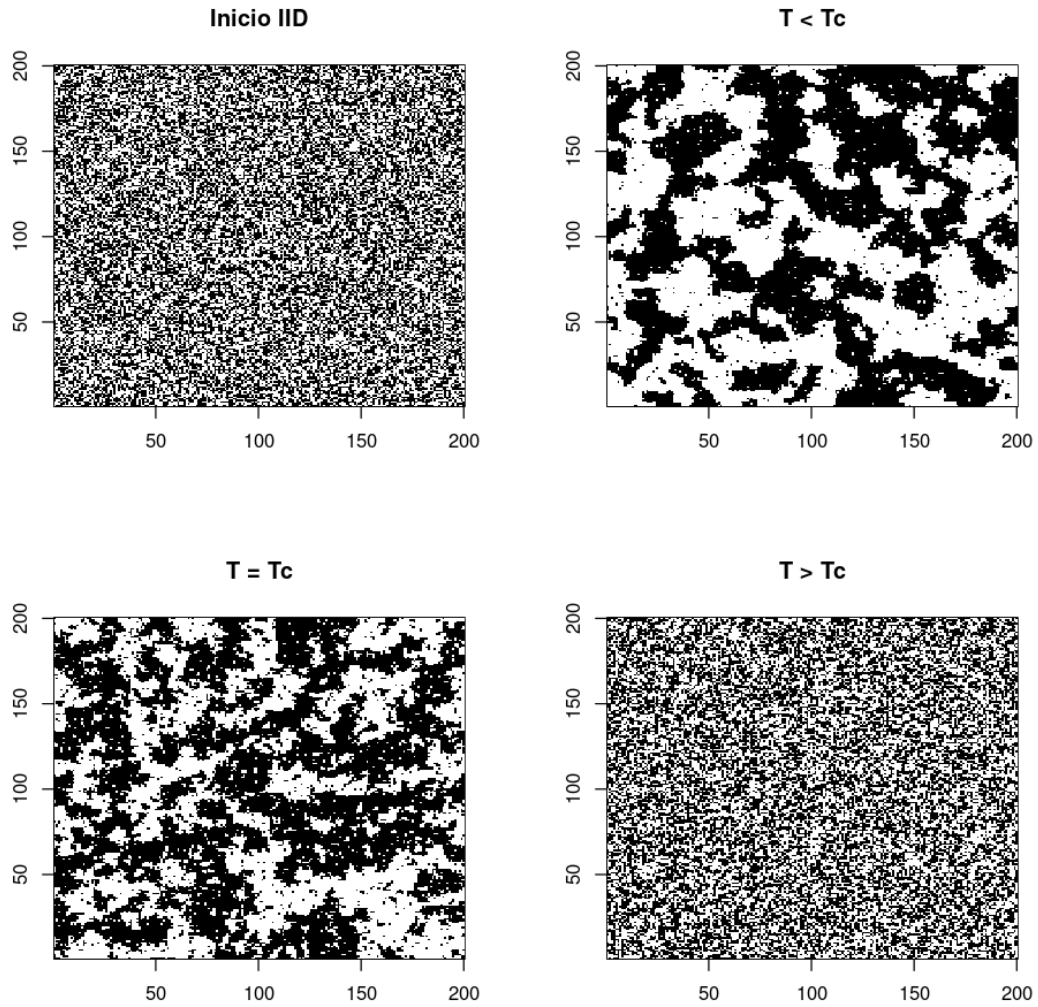


Figura 3.2: El modelo Ising con condiciones de frontera periódicas sobre una caja 200×200 a temperatura baja, crítica, y alta, respectivamente.

Conclusiones

En el capítulo 1 se hizo uso del poder del cálculo computacional para generar números con un comportamiento aleatorio llamados números pseudoaleatorios. En seguida se mostró algoritmos de simulación de variables aleatorias discreta o continua, específicamente los algoritmos método de la transformada inversa y método de aceptación/rechazo. En el capítulo 2 se estableció las técnicas estadísticas que permitirán estimar un parámetro característico de una población de variables aleatorias mediante el estudio de una muestra. También, se desarrolla el método de integración por Monte Carlo. Finalmente se mostró que el método de Monte Carlo con cadenas de Markov es flexible y manejable al momento de simular valores según una distribución arbitraria, así como se presentó al modelo de Ising consiguiendo hacer una simulación de la transición de fase en dimensión $d = 2$.

Apéndice

Programa 1 (El método de integración por Monte Carlo. Ver figura 2.2).

```
#####  
##### Integral de e^x #####  
#####  
## \int_{-2}^2 \exp(x)dx ##  
#####  
  
integral1<- function(k){  
## k: Numero de iteraciones  
  
# repite "k" veces el "0", luego los concatena en un vector "s"  
s <- c(rep(0,each=k)); x <- 1:k  
  
s[1] = 4*exp(runif(1,-2,2))  
for (i in 2:k) {  
# genera valores entre -2 y 2 segun distribución uniforme  
s[i] <- s[i-1]+4*exp(runif(1,-2,2));  
}  
  
int <- s/x  
jpeg(filename="inte-1a.jpeg", width = 30, height = 21, res= 72, units = "cm")  
plot(x,int)
```

```

abline(a=0,b=10**4,h=7.25372081569404,col='red')
dev.off()          # Cierre del archivo
integral<-int[k]
integral
}

k=10**3
Inew = integral1(k)
print( c("montecarlo =",k,Inew) )
#Iteo = exp(2)-exp(-2)
#print( c("teórico =","/infty",Iteo) )

```

Programa 2 (El método de Monte Carlo con intervalo de confianza. Ver figura 2.3.).

```
#####
##### Integral de e^x #####
#####
## \int_{-2}^2 \exp(x)dx ##
#####

integral1<- function(error){
## error: cota de error para el estimador

# inicialización
x <- c(rep(0,each=2)); s <- c(rep(0,each=2)); n <- 1:2
x[1] = 4*exp(runif(1,-2,2)); x[2] = 0.5*(4*exp(runif(1,-2,2))+x[1]);
s[1] = 0; s[2] = 2*(x[2]-x[1])**2; i=2;

alpha = 0.05; z = -qnorm(alpha/2)
while ( 2.0*z*sqrt(1.0*s[i]/i)>= error ) {
x <- unlist( list(x,c(0)) )
s <- unlist( list( s, c(0)) ); n <- unlist( list(n,c(i+1)) )
# genera valores entre -2 y 2 segun distribución uniforme
x[i+1] <- x[i]+(1./(i+1))*(4*exp(runif(1,-2,2))-x[i]);
s[i+1] <- (1.-1./i)*s[i]+(i+1.0)*(x[i+1]-x[i])**2;
i <- i+1;
}

# Grafico
jpeg(filename="inte-1b.jpeg", width = 30, height = 21, res= 72, units = "cm")
plot(n,x); abline(a=0,b=10,h=7.25372081569404,col='red'); dev.off()

# Intervalo de Confianza
print(c("nro de iteraciones = ",i))
```

```

integral<-x[i]; print(c("integral = ",integral))
print(c("varianza = ",s[i])); alpha = 0.05; z = -qnorm(alpha/2)
I = c( integral -z*sqrt(s[i]/i), integral +z*sqrt(s[i]/i) )
print(c("intervalo al 95% =",I))
print( c("teórico =",exp(2)-exp(-2)) ) #Iteo = exp(2)-exp(-2)
return(integral)
}

error=0.01
Inew = integral1(error)

```

Programa 3 (El método de Monte Carlo con intervalo de confianza. Ver figura 2.4.).

```
#####
### Regla de las tres sigmas ###
##### normal estándar #####
#####
## \int_{-3}^3 \exp(-x^2/2)dx ##
#####

integral1<- function(error){
## error: cota de error para el estimador

# inicialización
x <- c(rep(0,each=2)); s <- c(rep(0,each=2)); n <- 1:2
x[1] = 6*exp(-0.5*runif(1,-3,3)**2); x[2] = 0.5*(6*exp(-0.5*runif(1,-3,3)**2)+x[1]);
s[1] = 0; s[2] = 2*(x[2]-x[1])**2; i=2;

alpha = 0.05; z = -qnorm(alpha/2)
while ( 2.0*z*sqrt(1.0*s[i]/i)>= error ) {
x <- unlist( list(x,c(0)) )
s <- unlist( list( s, c(0)) ); n <- unlist( list(n,c(i+1)) )
# genera un valor entre -3 y 3 segun distribución uniforme
x[i+1] <- x[i]+(1./(i+1))*(6*exp(-0.5*runif(1,-3,3)**2)-x[i]);
s[i+1] <- (1.-1./i)*s[i]+(i+1.0)*(x[i+1]-x[i])**2;
i <- i+1;
}

### grafico
jpeg(filename="inte-2b.jpeg", width = 30, height = 21, res= 72, units = "cm")
plot(n,x); abline(a=0,b=10,h=2.499108389807114,col='red'); dev.off()

### intervalo de confianza
```

```

print(c("nro de iteraciones = ",i))
integral<-x[i]; print(c("integral = ",integral))
print(c("varianza = ",s[i])); alpha = 0.05; z = -qnorm(alpha/2)
I = c( integral -z*sqrt(s[i]/i), integral +z*sqrt(s[i]/i) )
print(c("intervalo al 95% =",I))
print( c("teórico =",sqrt(2*pi)*0.997) ) #Iteo = sqrt(2*pi)*0.997
return(integral)
}

error=0.02
Inew = integral1(error)

```

Aquí se presenta el código en R usado para generar la Figura 3.2.

```

##### Modelo de Ising en  $Z^2$  #####
# NxN: tamaño del bloque en  $Z^2$ 
# B: configuración en el bloque
# k: número de iteraciones
# be:  $\beta = 1/T$ 

# configuración inicial iid
init <- function(N){
return( array( sample(c(-1,1),size=N*2,replace=TRUE), dim=c(N,N)) )
}

#B <- init(N);
#image(1:500,1:500, B, col=c("white","black"), ylab="", xlab="");

# Gibbs Sampler
ising <- function(be,B,N,k){

```

```

#B <- init(N);
for (i in 0:k){
## elegir la posicion a cambiar
ri <- sample(N,1); rj <- sample(N,1);
## medida condicional
S = B[1+((N-1)+ri+1)%N,rj]+B[1+((N-1)+ri-1)%N,rj];
S = S + B[ri,1+((N-1)+rj+1)%N]+B[ri,1+((N-1)+rj-1)%N];
p <- 1./(1+exp(-2*be*S));
## determinar si se actualiza B[ri,rj]=+1 o B[ri,rj]=-1
u <- runif(1);
if (u < p){
B[ri,rj] <- +1;
}
if (u >= p){
B[ri,rj] <- -1;
}
}
return(B)
}

```

```

##### Transicion de fase en el Modelo de Ising en  $Z^2$  #####
N=200; k=50*N*N; Tc=2.0/log(1+sqrt(2)); ##
png("gibbs-measure.png", width=8, height=8, units="in", res=120)
par(mfrow=c(2,2))
B=init(N);
image(1:N,1:N,B,col=c("white","black"), ylab="", xlab="", main="Inicio IID")
temp = 0.8*Tc; be = 1/temp;
B1 <- ising(be,B,N,k);

```



```
image(1:N,1:N,B1,col=c("white","black"), ylab="", xlab="", main="T < Tc")
temp = Tc; be = 1/temp;
B2 <- ising(be,B,N,k);
image(1:N,1:N,B2,col=c("white","black"), ylab="", xlab="", main="T = Tc")
temp = 4.0*Tc; be = 1/temp;
B3 <- ising(be,B,N,k);
image(1:N,1:N,B3,col=c("white","black"), ylab="", xlab="", main="T > Tc")
dev.off()
```

Bibliografía

- [1] Giuseppe Dodge, Yadolah; Melfi. *Premiers pas en simulation*. Springer, Paris, first edition, 2008.
- [2] Jacques Franchi. *Processus aléatoires à temps discret*. Ellipses Édition Marketing S.A., Paris, first edition, 2013.
- [3] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.
- [4] James E. Gentle. *Random number generation and Monte Carlo methods*. Statistics and Computing. Springer, New York, second edition, 2003.
- [5] David A Levin, Yuval Peres, and EL Wilmer. *Markov chains and mixing times: Second Edition*. American Mathematical Society, Providence, 2017.
- [6] Sheldon M. Ross. *Simulation*. Elsevier/Academic Press, Amsterdam, 2013. Fifth edition.
- [7] GH Wannier and HA Kramers. Statistics of the two-dimensional ferromagnet. part I. *Phys. Rev*, 60:252–262, 1941.