

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE CIENCIAS



ESCUELA PROFESIONAL DE MATEMÁTICA

TRABAJO MONOGRÁFICO EN

Una breve introducción a la simulación

Alumno:

Camarena Pérez, Victor Daniel

Asesor:

Espejo Delzo, Juan Carlos

Lima, Perú
2017

Resumen

Nuestro objetivo 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.

Índice general

Índice General	0
Introducción	1
1. Generación de números y variables aleatorias	3
1.1. Generación de números aleatorios	3
1.2. Generación de variables aleatorias discretas	6
1.2.1. El método de la transformada inversa	6
1.2.2. La técnica de aceptación y rechazo	9
1.3. Generación de variables aleatorias continuas	11
1.3.1. El algoritmo de la transformada inversa	11
1.3.2. El método de rechazo (caso continuo)	13
1.3.3. Métodos para generar variables aleatorias normales	15
2. Análisis estadístico de datos simulados	22
2.1. Estimación puntual	22
2.2. Estimación por intervalos de confianza	27
2.3. Integración por Monte Carlo	30
Apéndice	35
Cálculo de probabilidades	35
Programas usados en el documento	48
Bibliografía	54

Introducción

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.

El ejemplo anterior muestra como la teoría de la simulación provee de un método alternativo para resolver un problema matemático, el cálculo del valor de una 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 se usa cada vez con más frecuencia.

El presente trabajo sigue el desarrollo realizado en el libro [6], y consta de dos capítulos y un apéndice. En el capítulo 1 se hace uso del poder del cálculo computacional para generar números con un comportamiento aleatorio, mas como no son perfectamente aleatorios son llamados números pseudoaleatorios. En seguida se muestra como a partir de la generación de números aleatorios

(pseudoaleatorios) se pueden generar valores de una variable aleatoria discreta o continua, además se hace un estudio de dos algoritmos que permiten generar valores de una variable aleatoria normal. 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, es decir, resolvemos el problema de estimación de dicho parámetro. A continuación, se desarrolla el método de integración por Monte Carlo como un caso particular del problema de estimación ya estudiado. El apéndice consta de dos partes. La primera parte del apéndice trata sobre el cálculo de probabilidades con la finalidad de mostrar la prueba del Teorema central del límite, el cual es clave para resolver el problema de estimación que abordamos en el presente trabajo. La segunda parte del apéndice es el listado de los programas, en R, usados para generar las gráficas mostradas en el presente trabajo.

Por último, recalcar que actualmente los métodos desarrollados por la teoría de la simulación son aplicados para resolver diversos problemas que aparecen en física, economía, ingeniería y otras áreas, obteniéndose muy buenos resultados.

Capítulo 1

Generación de números y variables aleatorias

1.1. 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

$$x_n = (ax_{n-1}) \text{ mód } m, \quad (1.1)$$

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 [2], 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

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

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.

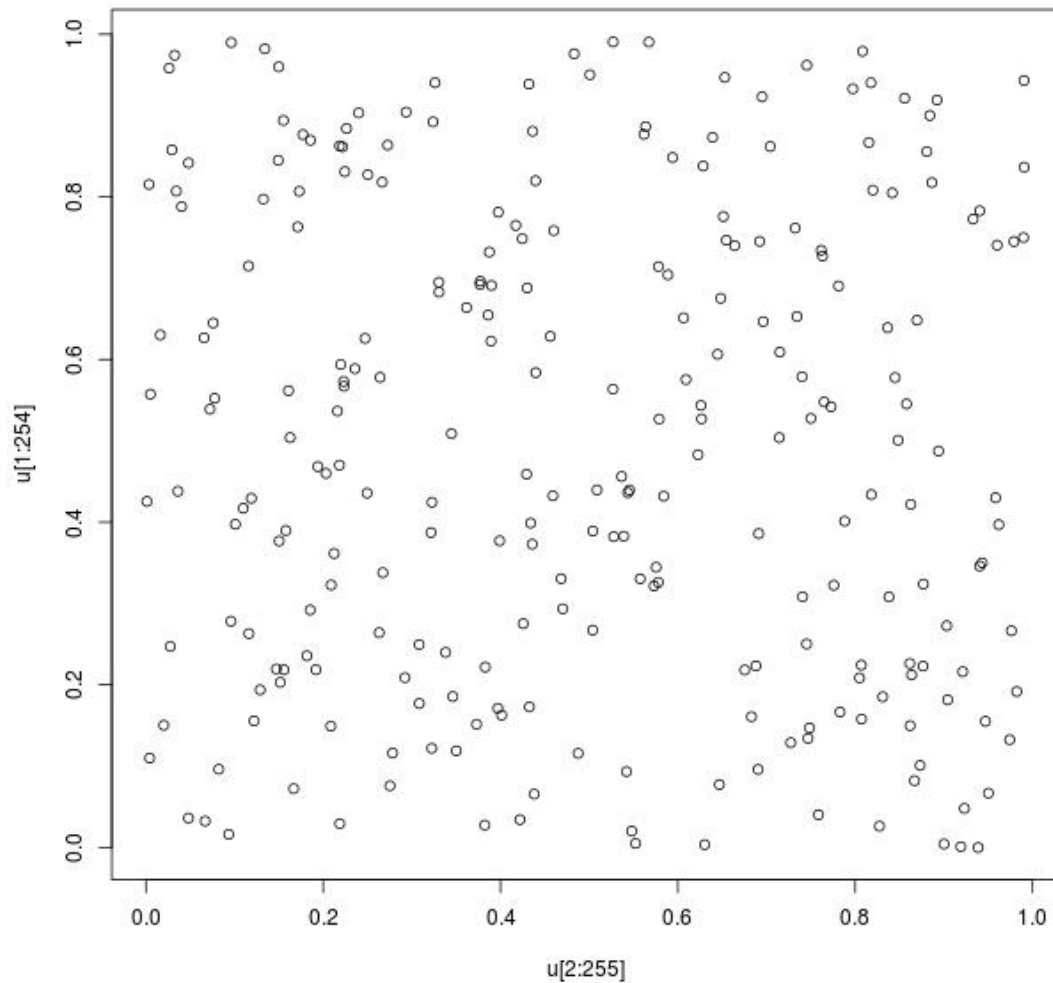


Figura 1.1: El método congruencial mixto
Ver programa 1 del apéndice.

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.

1.2. Generación de variables aleatorias discretas

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.

1.2.1. El método de la transformada inversa

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

$$\mathbb{P}(X = x_j) = p_j, \quad j = 0, 1, \dots, \sum_j p_j = 1. \quad (1.2)$$

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

$$X = x_j \quad \text{si} \quad \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i, \quad j = 0, 1, 2, \dots \quad (1.3)$$

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

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

²Ver definición 13 del apéndice.

y por lo tanto, X tiene la distribución deseada.

Observación 1. 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 (1.3) se convierte en

$$X = x_j, \quad \text{si } F(x_{j-1}) \leq U < F(x_j). \quad (1.4)$$

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 2. 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 1 (Generación de una variable aleatoria geométrica). Sea X una variable aleatoria geométrica con parámetro p , es decir, su función de masa de probabilidad es

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

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 (1.3),

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

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 .

Ejemplo 2 (Generación de una variable aleatoria Poisson). Sea X una variable aleatoria de Poisson con media λ , entonces

$$p_i = \mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \dots$$

Esta función de masa de probabilidad cumple la siguiente recursión

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

Se considera las notaciones: $pr = p_i$ la probabilidad de que X sea igual a i y $F = F(i)$ la probabilidad de que X sea menor o igual a i . Así el algoritmo de la transformada inversa para generar una variable aleatoria de Poisson se puede escribir

1. Generar un número aleatorio U .
2. $i = 0$, $pr = e^{-\lambda}$, $F = pr$.
3. Si $U < F$, hacer $X = i$ y terminar. Sino, hacer $pr = \frac{\lambda}{i+1} pr$, $F = F + pr$, $i = i + 1$.
4. Ir al paso 3.

Observación 3. El algoritmo anterior verifica en forma sucesiva si el valor de Poisson es 0, luego si es 1, luego 2, etc. Así, el número de comparaciones necesarias será 1 más que el valor generado de Poisson. Por lo tanto, en promedio, lo anterior necesitará realizar “1 más el valor esperado de la variable aleatoria de Poisson” de búsquedas, es decir, $1 + \lambda$ búsquedas.

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

1.2.2. La técnica de aceptación y rechazo

Suponga que tenemos un método eficiente para simular una variable aleatoria con función de masa de probabilidad $\{q_j, j \geq 0\}$. Podemos emplearlo como 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

$$\frac{p_j}{q_j} \leq c \quad \forall j \geq 0 \text{ con } p_j > 0. \quad (1.5)$$

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 2 (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 1. Sean $\{p_j, j \geq 0\}$ y $\{q_j, j \geq 0\}$ dos de masa de probabilidad y sea $c > 0$ de modo que se cumple la desigualdad (1.5). 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:

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

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,

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

□

Observación 4.

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 (1.5) 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 3. 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

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

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 .

1.3. Generación de variables aleatorias continuas

Ya en la sección anterior se mostró como generar valores (pseudoaleatorios) de una variable aleatoria discreta, ahora, siguiendo una idea similar trabajaremos el caso de una variable aleatoria continua.

1.3.1. El algoritmo de la transformada inversa

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

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

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 2. 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 3 (El método de la transformada inversa).

1. Generar un número aleatorio U .
2. Calcular $X = F^{-1}(U)$.

Ejemplo 4. 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

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

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

$$F(x) = \int_0^x \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!} dt,$$

y de la cual no es posible dar una expresión con forma cerrada de su inversa. Por otro lado tenemos que una variable aleatoria gamma X con parámetros (n, λ) se puede considerar como la suma de n exponenciales independientes, cada una con razón λ , las cuales ya sabemos generar. Entonces, podemos obtener una variable aleatoria gamma (n, λ) si generamos n números aleatorios U_1, \dots, U_n y hacemos

$$X = -\frac{1}{\lambda} \log U_1 - \dots - \frac{1}{\lambda} \log U_n = -\frac{1}{\lambda} \log(U_1 \cdots U_n).$$

1.3.2. El método de rechazo (caso continuo)

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 4 (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 3. 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 5.

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 6. 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,

$$c = \max \frac{f(x)}{g(x)} = \frac{3K}{2} \left(\frac{3}{2}\right)^{1/2} e^{-1/2},$$

$$= \frac{3^{3/2}}{2\pi e^{1/2}}, \quad \text{pues } K = \frac{2}{\sqrt{\pi}},$$

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

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

1.3.3. Métodos para generar variables aleatorias normales

El método de rechazo

Un primer algoritmo para generar una variable aleatoria normal mediante el método del rechazo. Para generar sea Z una variable aleatoria normal estándar, primero observe que el valor absoluto de Z tiene función de densidad de probabilidad

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad x > 0 \tag{1.7}$$

Así, primero generamos un valor de la variable aleatoria $|Z|$ a partir de la función de densidad anterior, valiéndonos del método de rechazo en el que se considera g como la función de densidad exponencial con media 1 (es decir, $g(x) = e^{-x}$, $x > 0$). Ahora,

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

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

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

Luego

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

y esto implica que se puede generar el valor absoluto de una variable aleatoria normal estándar como sigue:

1. Generar Y , una variable aleatoria exponencial con razón 1.
2. Generar un número aleatorio U .
3. Si $U \leq \exp\{-(Y-1)^2/2\}$, hacer $X = Y$. En caso contrario, regresar al paso 1.

Una vez simulada una variable aleatoria $X = |Z|$ con función de densidad como en la ecuación (1.7), podemos obtener un valor de la normal estándar Z al hacer que sea igualmente probable que Z tome X o $-X$. Por tanto, el paso 3 queda modificado como sigue:

3. Si $U \leq \exp\{-(Y-1)^2/2\}$ hacer $X = Y$, luego, generar un número aleatorio V y asignar Z de la siguiente manera

$$Z = \begin{cases} X & V \leq 0,5 \\ -X & V > 0,5 \end{cases}$$

En caso contrario, regresar al paso 1.

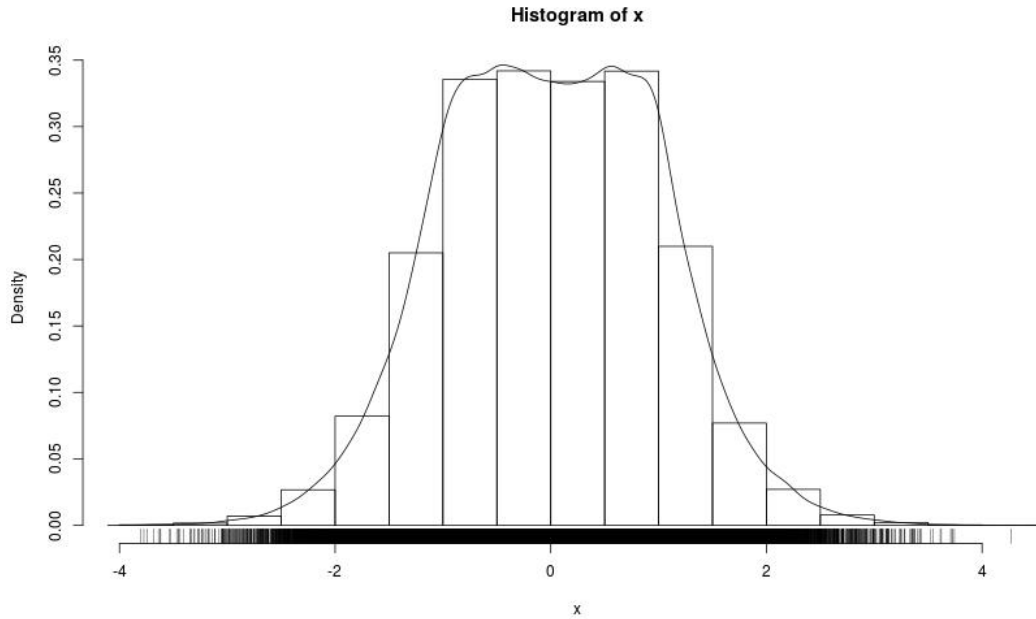


Figura 1.2: El método del rechazo.
Ver programa 2 del apéndice.

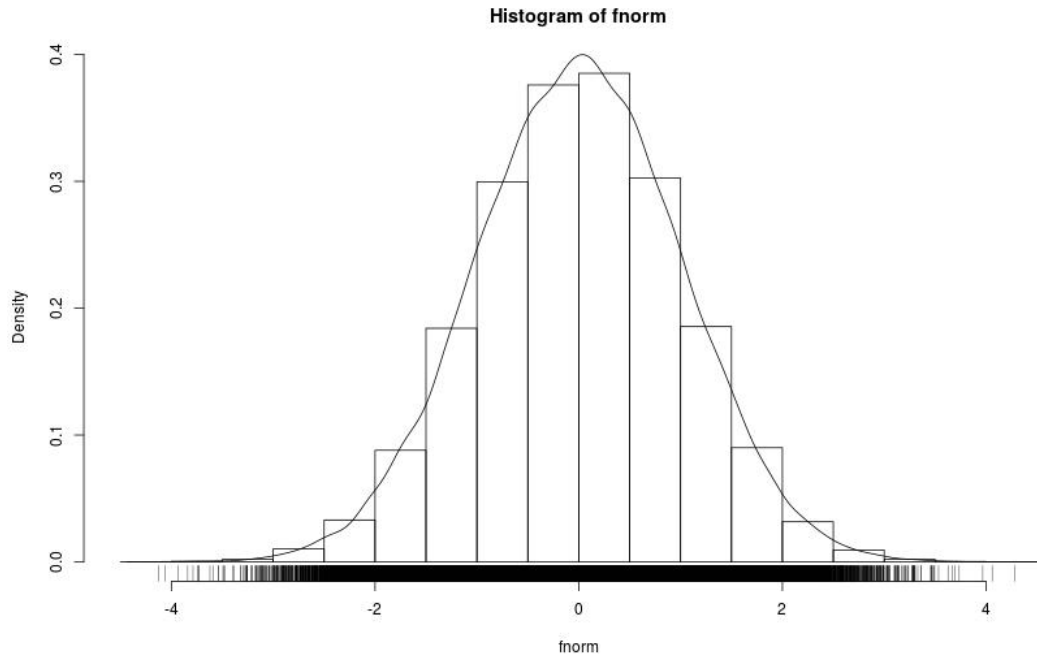


Figura 1.3: Generador, por defecto en R, de una normal.
Ver programa [2](#) del apéndice.

En la figura [1.2](#) se observa el histograma de frecuencias de una muestra de 50000 valores generados por este método, note que tiene la forma de la función de densidad de una normal estándar. La figura [1.3](#) muestra el histograma de frecuencias para una cantidad de 50000 valores de una normal generados mediante la sentencia de R `rnorm(50000, 0, 1)`, que es el generador por defecto en R.

Una comparación de estas gráficas nos indican si nuestro algoritmo produce un valor de una normal estándar, pues es de esperar que al generar una gran cantidad de estos valores y graficar su histograma de frecuencias nos resulte algo similar a la figura [1.3](#). En todo caso, la aparición de una loma en torno a la media en la figura [1.2](#), muestra que la calidad de un valor generado por nuestro algoritmo no es tan bueno como el generador por defecto en R.

El método polar

Sean X y Y variables aleatorias normales unitarias independientes y sean R y Θ las coordenadas polares del vector (X, Y) . Es decir

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

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

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

$$f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad (1.8)$$

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

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

Como es fácil ver que el jacobiano de esta transformación, $(d, \theta) = (d(x, y), \theta(x, y))$, es igual a 2, la ecuación (1.8) junto con el teorema del cambio de variable⁴ implican que la función de densidad conjunta de R^2 y Θ está dada por

$$f(d, \theta) = \frac{1}{2} \frac{1}{2\pi} e^{-d/2}, \quad 0 < d < \infty, \quad 0 < \theta < 2\pi \quad (1.9)$$

Calculando se obtiene que las densidades marginales de R^2 , denotada $f(d)$, y Θ , denotada $f(\theta)$, son

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

es decir,

$$\begin{aligned} R^2 & \text{ es exponencial con media } 2 \\ \Theta & \text{ es uniforme sobre }]0, 2\pi[\end{aligned}$$

Además, como la densidad conjunta dada en por la ecuación (1.9) es igual al producto de las densidades de R^2 y Θ , $f(d)$ y $f(\theta)$, se tiene que

$$R^2 \text{ y } \Theta \text{ son independientes}$$

Ahora podemos generar un par de variables aleatorias normales estándar independientes X e Y , generando primero sus coordenadas polares $(R^2, \Theta) = (-2 \log U_1, 2\pi U_2)$, donde U_1 y U_2 son números aleatorios; y luego transformarlas a coordenadas rectangulares $(X, Y) = (R \cos \Theta, R \sin \Theta)$. Tal como se detalla a continuación.

⁴Sean $f(x, y)$ y $g(u, v)$ dos funciones reales, sea la transformación $T = (T_1, T_2)$ verificando la igualdad $(u, v) = (T_1(x, y), T_2(x, y))$. Entonces se cumple

$$g(u, v) = \frac{1}{JT(x, y)} f(T^{-1}(u, v)),$$

donde $JT(u, v)$ es el valor absoluto del determinante de la derivada de T , es decir,

$$\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix}.$$

Teorema 4. Sean U_1 y U_2 dos variables aleatorias uniformes e independientes sobre el intervalo $]0, 1[$. Entonces las variables

$$\begin{aligned} X &= \sqrt{-2 \log U_1} \cos(2\pi U_2) \\ Y &= \sqrt{-2 \log U_1} \sin(2\pi U_2) \end{aligned} \quad (1.10)$$

son dos variables aleatorias normales unitarias e independientes.

Esto se permite establecer el siguiente algoritmo:

Algoritmo 5 (El método de Box-Muller).

1. Generar números aleatorios U_1 y U_2 .
2. Hacer $R^2 = -2 \log(U_1)$ y $\Theta = 2\pi U_2$.
3. Asignar

$$\begin{aligned} X &= \sqrt{R^2} \cos(\Theta), \\ Y &= \sqrt{R^2} \sin(\Theta). \end{aligned}$$

Sin embargo, el uso de las transformaciones de Box-Muller (1.10) para generar un par de normales unitarias independientes no es eficiente desde el punto de vista computacional, y la razón está en la necesidad de calcular las funciones trigonométricas seno y coseno. Por ello, una forma de mejorar la eficiencia es mediante un cálculo indirecto del seno y el coseno de un ángulo aleatorio. Si generamos números aleatorios U_1 , y U_2 y hacemos

$$(V_1, V_2) = (2U_1 - 1, 2U_2 - 1),$$

entonces se puede probar que (V_1, V_2) está uniformemente distribuido en el cuadrado de área 4 con centro en $(0, 0)$.

Luego, si generamos de manera continua tales pares (V_1, V_2) hasta obtener uno que esté dentro del círculo de radio 1 con centro en $(0, 0)$ se tiene que tal par (V_1, V_2) está uniformemente distribuido en el círculo. Sean R y Θ las coordenadas polares de este par, entonces se logra verificar que R^2 y Θ son independientes, que R^2 está uniformemente distribuida en $]0, 1[$ y que Θ está uniformemente distribuida en $]0, 2\pi[$. Así, Θ es un ángulo aleatorio, luego podemos calcular el seno y el coseno de este un ángulo aleatorio Θ haciendo

$$\begin{aligned} \cos \Theta &= \frac{V_1}{(V_1^2 + V_2^2)^{1/2}}, \\ \sin \Theta &= \frac{V_2}{(V_1^2 + V_2^2)^{1/2}}. \end{aligned}$$

Además, como $R^2 = V_1^2 + V_2^2$ está uniformemente distribuido en $]0, 1[$ y es independiente del ángulo aleatorio Θ , podemos tomarlo como el número aleatorio U_1 necesario en las ecuaciones (1.10). Así si $U_1 = R^2$ y $\Theta = 2\pi U_2$, las ecuaciones de Box-Muller quedan como siguen

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

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

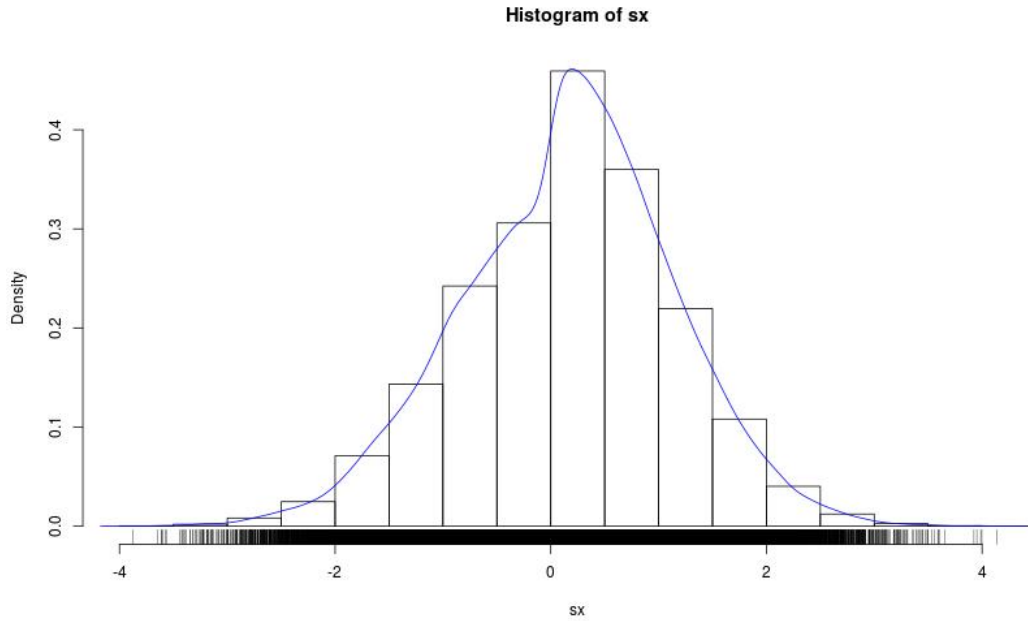


Figura 1.4: El método polar. Densidad marginal en el eje X .
Ver programa 3 del apéndice.

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

Algoritmo 6 (El método polar).

1. Generar números aleatorios U_1 y U_2 .
2. Hacer $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$.
3. Si $S \leq 1$ asignar X e Y como sigue

$$X = \left(\frac{-2 \log S}{S} \right)^{1/2} V_1, \quad Y = \left(\frac{-2 \log S}{S} \right)^{1/2} V_2.$$

En caso contrario regresar al paso 1.

Observación 6. La probabilidad de que un punto aleatorio del cuadrado esté dentro del círculo es igual a $\pi/4$ (el área del círculo dividido entre el área del cuadrado). Así, el número de puntos aleatorios del cuadrado generados hasta que obtenemos uno que esté dentro del círculo sigue una distribución geométrica con parámetro $\pi/4$, y entonces el método necesitará, en promedio, $4/\pi = 1,273$ iteraciones para generar dos normales unitarias independientes.

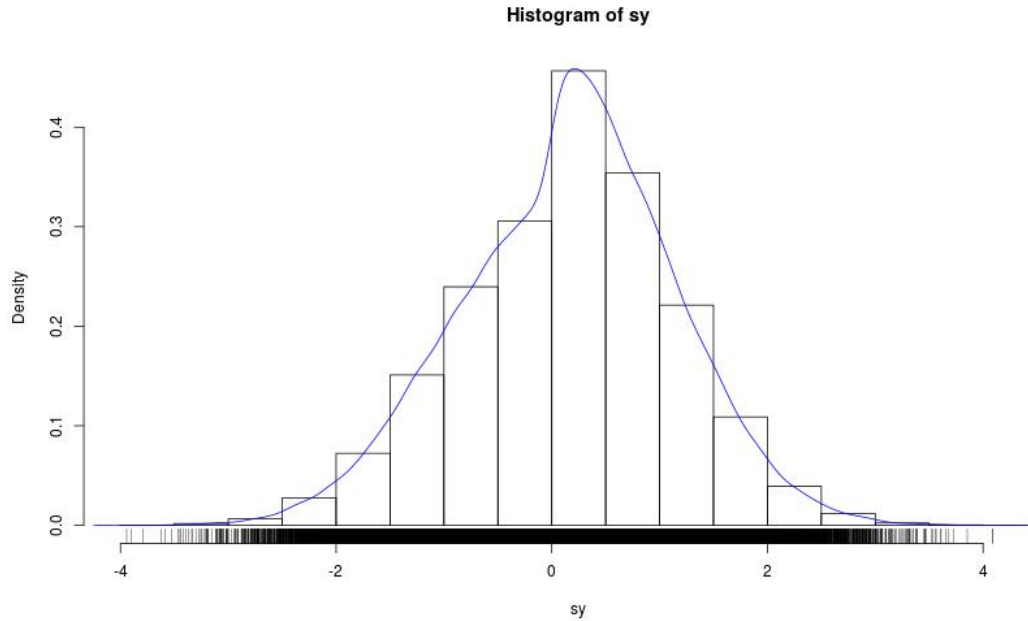


Figura 1.5: El método polar. Densidad marginal en el eje Y .
Ver programa 3 del apéndice.

En la figura 1.4 se observa el histograma de frecuencias marginal a X de una muestra de 50000 valores generados con el método polar, note que tiene la forma de la función de densidad de una normal estándar. La gráfica del histograma de frecuencias marginal a Y es muy similar a la de X , esto se aprecia en la figura 1.5. Así, se comprueba que las variables X e Y son independientes.

Recordando la comparación hecha entre las figuras 1.2 y 1.3, y observando la figura 1.5, se puede de acotar que el algoritmo polar es más adecuado para la generación de valores de una normal estándar. Este algoritmo tiene dos ventajas, primero es de menor costo computacional y segundo genera dos valores correspondientes a dos normales unitarias independientes. Sin embargo, este generador polar, aún sigue siendo menos bueno que el generador por defecto en R.

Capítulo 2

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 n valores, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ sirve entonces como estimador, o aproximador, de θ .

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.

2.1. Estimación puntual

Definición 1. 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 $\mathbb{E}[\theta_n] = \theta$.

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.

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 2. 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 7. 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 3. 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,

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

Así, la media muestral, \bar{X}_n , es una variable aleatoria con media θ y varianza $\frac{\sigma^2}{n}$. La desigualdad de Chebyshev³ 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 4. 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 1. La varianza muestral es un estimador insesgado, es decir,

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

³Ver proposición 6 del apéndice.

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 $n \geq 1$.

Con estas consideraciones previas se obtiene el siguiente algoritmo.

Algoritmo 7 (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

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

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

3. Mientras $\frac{S_n}{\sqrt{n}} > \epsilon$ 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. La estimación de θ está dada por \bar{X}_n .

2.2. 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 5. 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 8. El intervalo I_n es aleatorio mientras que θ es fijo y determinado.

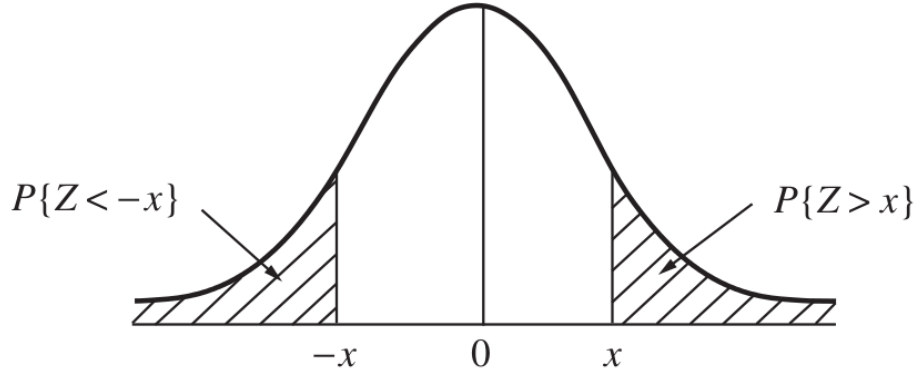


Figura 2.1: Densidad normal estándar

Fuente: Ver figura 8.1 de [6].

Teorema 5 (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 14, 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 9. 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 8 (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

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

3. Mientras $2z_{\alpha/2} \frac{S_n}{\sqrt{n}} > \epsilon$ 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. El intervalo de confianza del $100(1 - \alpha)\%$, con longitud menor que ϵ , es

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

2.3. Integración 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 9 (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 10 (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

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

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

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

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

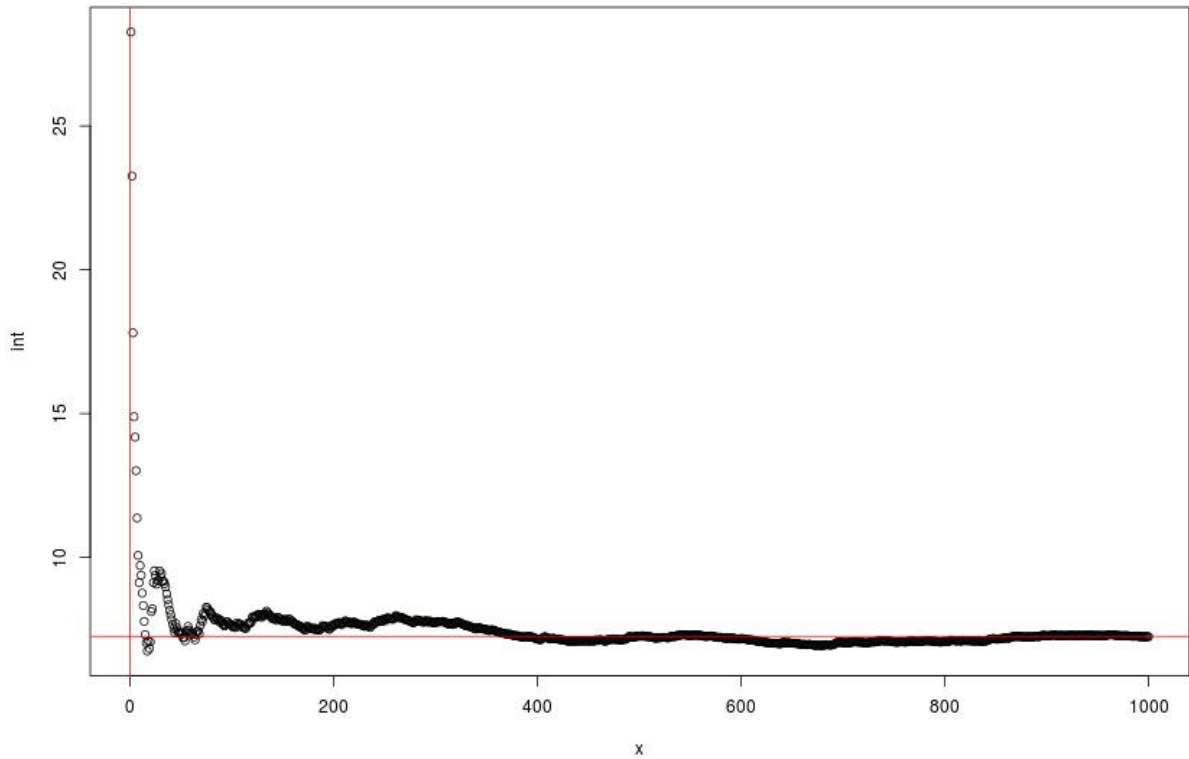


Figura 2.2: Integral de $\exp(x)$ para $n = 10^4$.
Ver programa 4 del apéndice.

Ejemplo 7. 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 (ver programa 5), 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 %.

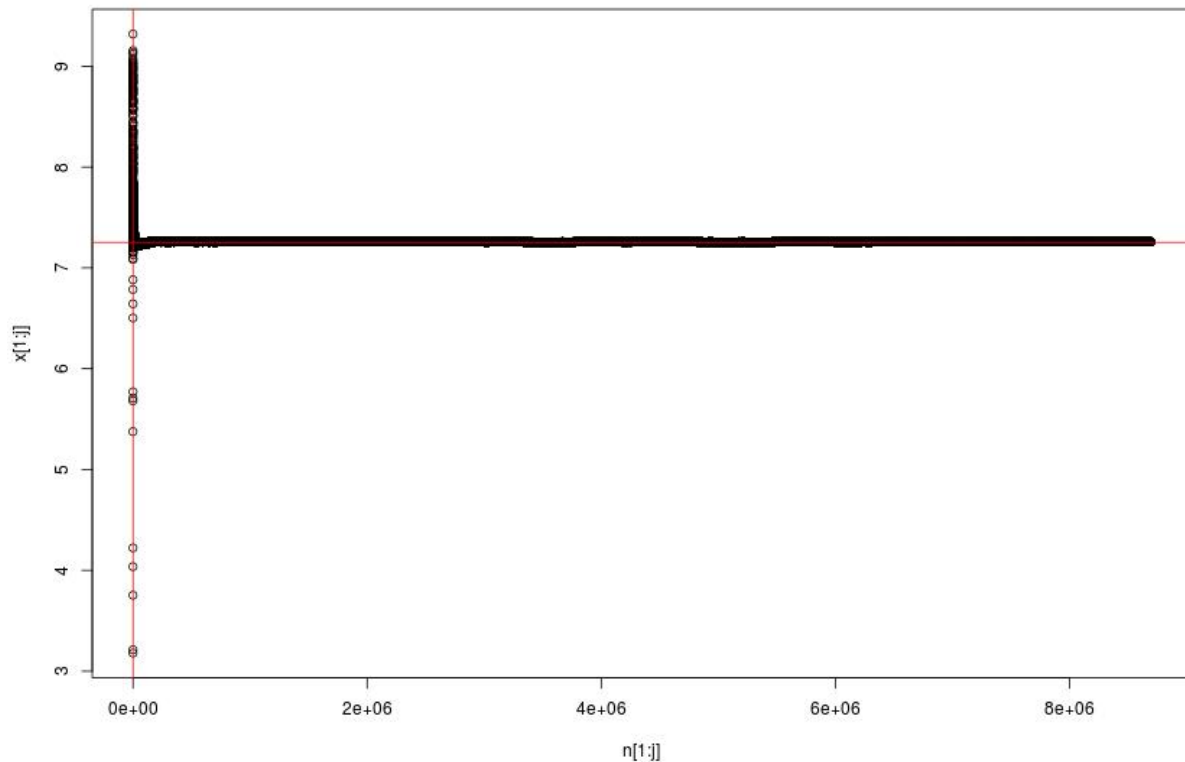


Figura 2.3: Integral de $\exp(x)$.
Ver programa 5 del apéndice.

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 8 (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 6), para $n = 167964$ iteraciones, nos da que la aproximación es

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

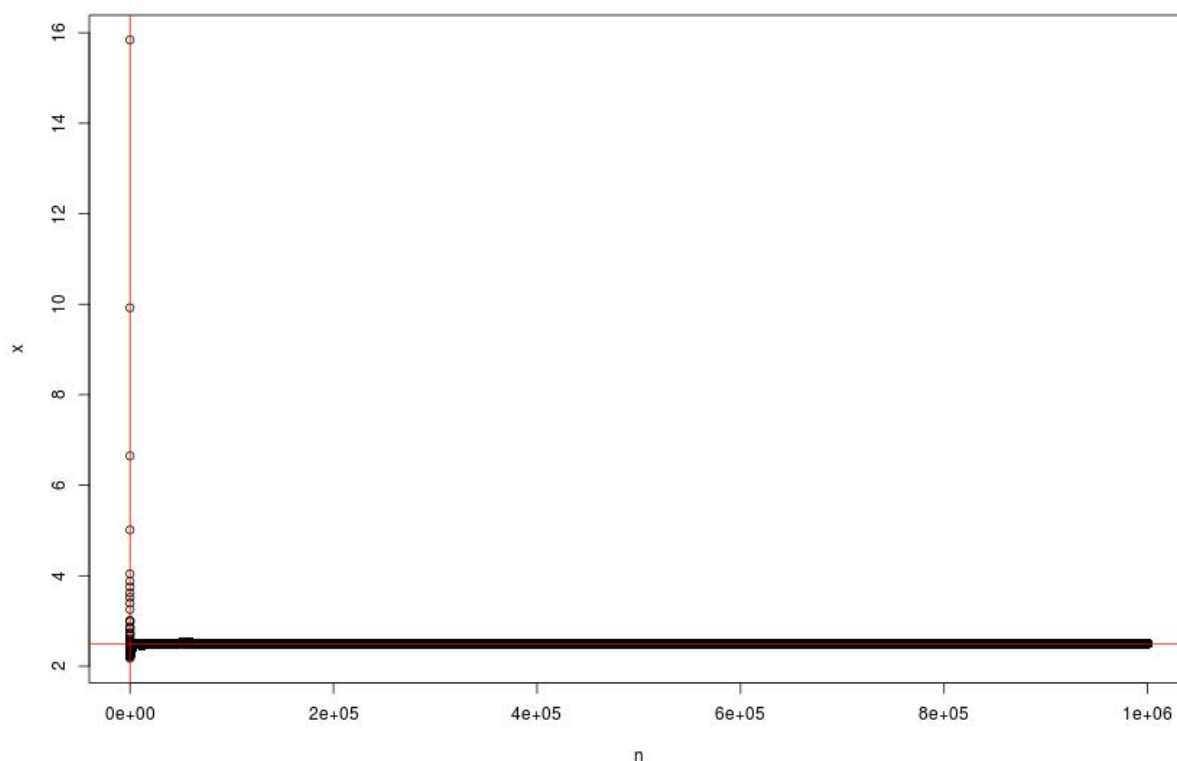


Figura 2.4: Integral de $\exp(-x^2/2)$.
Ver programa 6 del apéndice.

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 [2]. 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.

Apéndice

Cálculo de probabilidades

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 [3].

Probabilidad y condicionamiento

Definición 6. 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 10.

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 7. 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 2 (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. $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ 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 $\{B_n: n \geq 1\}$ (es decir, $B_{n+1} \subset B_n$).

Ejemplo 9. 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 8 (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}) .

Proposición 3. Sea un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una partición contable $\{C_j: j \geq 1\}$ de eventos no insignificantes: $\Omega = \bigcup_{j \geq 1} C_j$, $1 \leq j \neq k$ implica $C_j \cap C_k = \emptyset$. Entonces

1. (Fórmula de probabilidad total) para todo evento $A \in \mathcal{F}$ se cumple

$$\mathbb{P}(A) = \sum_{j \geq 1} \mathbb{P}(A|C_j)\mathbb{P}(C_j).$$

2. (Fórmula de Bayes) para todo evento $A \in \mathcal{F}$ no insignificante se cumple

$$\mathbb{P}(C_k/A) = \frac{\mathbb{P}(A/C_k)\mathbb{P}(C_k)}{\sum_{j \geq 1} \mathbb{P}(A/C_j)\mathbb{P}(C_j)}.$$

Definición 9. 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 10. 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 11 (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, V_2, \dots, V_n esta constituido por el σ -álgebra $\sigma\{V_1, V_2, \dots, V_n\} \subset \mathcal{F}$ que contiene todos los eventos de la forma $(V_1, V_2, \dots, V_n)^{-1}(E)$ 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 11. 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 12. 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} :

$$\mathbb{E}[V] = \int_{\Omega} V d\mathbb{P} \quad (2.8)$$

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 13 (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 4. 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 14. 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 15. 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

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

La covarianza de dos v.a. reales U, V cuadrado integrable es

$$\text{Cov}(U, V) = \mathbb{E}[(U - \mathbb{E}[U]) \times (V - \mathbb{E}[V])] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V].$$

En el lenguaje probabilístico es muy frecuente emplear el término *casi seguramente*, abreviado c.s., para referir que una propiedad o proposición tiene como dominio de validez a un evento casi seguro.

Proposición 5. La covarianza es bilineal, simétrica, positiva. La $\text{Var}(V)$ se anula si y solo si V es constante c.s., y para todo $\lambda \in \mathbb{R}$ se cumple la igualdad

$$\text{Var}(\lambda U + V) = \lambda^2 \text{Var}(U) + 2\lambda \text{Cov}(U, V) + \text{Var}(V).$$

Se cumple la desigualdad de Schwarz:

$$|\text{Cov}(U, V)| \leq \sigma(U)\sigma(V),$$

donde $\sigma(V) = \sqrt{\text{Var}(V)}$ es la desviación estándar de la v.a. V .

Corolario 1. El coeficiente de correlación lineal

$$\rho(U, V) = \frac{\text{Cov}(U, V)}{\sigma(U)\sigma(V)}$$

(definido para todo par U, V de v.a. reales en L^2 que no sean constantes casi seguramente) pertenece a $[-1, 1]$. Además, $\rho(U, V) = \pm 1$ si y solo si $U = aV + b$ (casi seguramente, para algunas a, b reales fijos).

Proposición 6 (Desigualdad de Markov, desigualdad de Chebyshev). Sea V una v.a. integrable y $f: [0, +\infty[\rightarrow [0, +\infty[$ una función creciente. Para cada $\epsilon > 0$ se cumple la desigualdad de Markov

$$\mathbb{P}(\|V\| \geq \epsilon) \leq \frac{1}{f(\epsilon)} \mathbb{E}[f(\|V\|)].$$

Si además V es una v.a. real cuadrado integrable, entonces para todo $\epsilon > 0$ se cumple la desigualdad de Chebyshev

$$\mathbb{P}(|V - \mathbb{E}[V]| \geq \epsilon \cdot \text{Var}(V)) \leq \frac{1}{\epsilon^2}.$$

Proposición 7. Sean U y V v.a. reales integrables. Si se cumple

1. $U \leq V$
2. $\mathbb{E}[U] = \mathbb{E}[V]$

entonces $U = V$ c.s..

Teorema 6 (Desigualdad de Jensen). Sea X una variable aleatoria integrable con valores en \mathbb{R}^d , y φ una función real convexa sobre \mathbb{R}^d . Entonces

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi \circ X].$$

Si $(E_i, \mathcal{F}_i, \mu_i)_{i=1,2}$ son dos espacios de medida, denotamos usualmente $\mathcal{F}_1 \otimes \mathcal{F}_2$ a la σ -álgebra producto generada por el par de proyecciones canónicas (π_1, π_2) , sobre $E_1 \otimes E_2$ y compuesta por los productos $A_1 \times A_2$ (para todo $A_i \in \mathcal{F}_i, i = 1, 2$), y $\mu_1 \otimes \mu_2$ la medida producto, que es una medida sobre $E_1 \times E_2$ tal que $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$ para todo $A_i \in \mathcal{F}_i$. Así, se construye el espacio de medida producto $(E_1 \times E_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$.

Teorema 7. *Dados $(E_i, \mathcal{F}_i, \mu_i)_{i=1,2}$ dos espacios de medida, y X una variable aleatoria real definida sobre $(E_1 \times E_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$, que se supone o positiva o integrable. Entonces se cumple*

$$\int_{E_1 \times E_2} X d\mu_1 \otimes \mu_2 = \int_{E_1} \left[\int_{E_2} X d\mu_2 \right] d\mu_1 = \int_{E_2} \left[\int_{E_1} X d\mu_1 \right] d\mu_2.$$

Variables aleatorias independientes

Definición 16. *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 8. *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 12.

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 9. *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.*

Variables aleatorias usuales

Sea X una v.a. definida sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con valores en \mathbb{R}^d .

Definición 17 (Variables aleatorias discretas). *Se dice que X tiene ley*

1. uniforme discreta, denotado $X \sim U(x_1, \dots, x_n)$, cuando X toma valores sobre $\{x_1, \dots, x_n\}$ y

$$U(x_1, \dots, x_n)(\{x_k\}) = \mathbb{P}(X = x_k) = \frac{1}{n}, \quad k = 1, \dots, n.$$

2. de Bernoulli de parámetro p ($0 \leq p \leq 1$), denotado $X \sim Be(p)$, cuando X toma valores sobre $\{0, 1\}$ y

$$Be(p)(\{1\}) = \mathbb{P}(X = 1) = p.$$

3. binomial de parámetros n y p (para todos $n \geq 1, 0 \leq p \leq 1$), denotado $X \sim B(n, p)$, cuando X toma valores sobre $\{0, 1, \dots, n\}$ y

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

4. geométrica de parámetro p ($0 \leq p \leq 1$), denotado $X \sim Ge(p)$, cuando X toma valores sobre \mathbb{N} y

$$Ge(p)(\{k\}) = \mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots$$

5. Poisson de parámetro λ ($\lambda > 0$), denotado $X \sim Po(\lambda)$, cuando X toma valores sobre $\{0, 1, 2, \dots\}$ y

$$Po(\lambda)(\{k\}) = \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Definición 18 (Variables aleatorias con función de densidad). *Se dice que X tiene ley*

1. uniforme sobre un abierto (o un cerrado) A de \mathbb{R}^d , denotado $X \sim U(A)$, cuando X toma valores sobre A y admite una función de densidad igual a $\frac{1}{\lambda(A)} 1_A(\cdot)$, donde λ es la medida de Lebesgue en \mathbb{R}^d , es decir,

$$U(A)(E) = \mathbb{P}(X \in E) = \int_E \frac{1}{\lambda(A)} 1_A(\cdot) d\lambda = \frac{\lambda(E \cap A)}{\lambda(A)},$$

para todo abierto $E \subset \mathbb{R}^d$.

2. exponencial de parámetro λ ($\lambda > 0$), denotado $X \sim Exp(\lambda)$, cuando X toma valores sobre \mathbb{R}_+ y admite una función de densidad igual a $h(t) = 1_{\mathbb{R}_+^*}(t) \lambda e^{-\lambda t}$, $t \in \mathbb{R}$, es decir,

$$Exp(\lambda)(E) = \mathbb{P}(X \in E) = \int_E 1_{\mathbb{R}_+^*}(t) \lambda e^{-\lambda t} \lambda(dt),$$

para todo abierto $E \subset \mathbb{R}$, donde λ es la medida de Lebesgue en \mathbb{R} .

3. normal de parámetros μ y σ^2 ($\mu \in \mathbb{R}, \sigma^2 > 0$), denotado $X \sim N(\mu, \sigma^2)$, cuando X toma valores sobre \mathbb{R} y admite una función de densidad igual a $h(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/(2\sigma^2)}$, $t \in \mathbb{R}$, es decir,

$$N(\mu, \sigma^2)(E) = \mathbb{P}(X \in E) = \int_E \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/(2\sigma^2)} \lambda(dt),$$

para todo abierto $E \subset \mathbb{R}$, donde λ es la medida de Lebesgue en \mathbb{R} . Cuando $\mu = 0$ y $\sigma^2 = 1$, X se dice normal estándar.

La función característica de una variable aleatoria

Definición 19 (Función característica). Sea V una v.a. real definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. La función característica de V es la función $\varphi_V: \mathbb{R} \rightarrow \mathbb{C}$ definida por

$$\varphi_V(t) = \mathbb{E} [e^{itV}],$$

donde $i = \sqrt{-1}$ es la unidad imaginaria.

Teorema 8 (Propiedades de la función característica).

1. La función característica es inyectiva: la ley de una variable aleatoria esta determinada por su función característica.
2. La función característica de una v.a. real $X \in L_p$, φ , es p -veces continuamente diferenciable con derivadas

$$\varphi^{(k)}(t) = \mathbb{E} [(iX)^k e^{itX}], \quad t \in \mathbb{R}, \quad k = 0, 1, \dots, p.$$

3. En particular, si X es cuadrado integrable entonces

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

donde $\rho(t) \rightarrow 0$ cuando $t \rightarrow 0$.

4. Además, si $\alpha \in \mathbb{R}$ no nulo entonces la función característica de αX esta dada por

$$\varphi_{\alpha X}(t) = \varphi_X(\alpha t), \quad t \in \mathbb{R}.$$

Corolario 2. Las v.a. reales V_1, \dots, V_n son independientes si y solo si se cumple

$$\varphi_{V_1+\dots+V_n}(t) = \varphi_{V_1}(t) \times \dots \times \varphi_{V_n}(t), \quad t \in \mathbb{R}.$$

Proposición 10. La función característica de una v.a. real N con media μ y varianza σ^2 esta dada por

$$\varphi_N(t) = e^{it\mu - \sigma^2 t^2/2}, \quad t \in \mathbb{R}.$$

Convergencia de variables aleatorias

Sea una sucesión de conjuntos $\{A_k: k \in \mathbb{N}\}$, se define su límite superior como

$$\limsup_{k \rightarrow \infty} A_k = \bigcap_k \bigcup_{l \geq k} A_k,$$

el conjunto de elementos que pertenecen a infinitos A_k . De igual manera, se define su límite inferior como

$$\liminf_k A_k = \bigcup_k \bigcap_{l \geq k} A_k,$$

el conjunto de elementos que pertenecen a todos los conjuntos A_k , excepto un número finito. De las definiciones se desprenden las siguientes propiedades:

1. $\left(\limsup_k A_k \right)^c = \liminf_k A_k^c$.
2. $1_{\left\{ \limsup_k A_k \right\}} = \limsup_k 1_{\{A_k\}}$.
3. $1_{\left\{ \liminf_k A_k \right\}} = \liminf_k 1_{\{A_k\}}$.
4. $\liminf_k A_k, \limsup_k A_k$ son eventos si los conjuntos $A_k, k \geq 1$, lo son.

Proposición 11 (Borel -Cantelli). *Sea la sucesión de eventos $\{A_k: k \in \mathbb{N}\}$ dentro de un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se cumple*

1. Si $\sum_k \mathbb{P}(A_k) < \infty$, entonces $\mathbb{P}\left(\limsup_k A_k\right) = 0$.
2. Si $\sum_k \mathbb{P}(A_k) = \infty$ y los $A_k, k \geq 1$, son independientes, entonces $\mathbb{P}\left(\limsup_k A_k\right) = 1$.

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 20. 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 12.

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 9.

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 13 (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 .

EL Teorema central del límite

Antes de enunciar el 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 10 (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 11 (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 13. *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 12 (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 $Z_k = \frac{X_1 + \cdots + X_k}{\sigma\sqrt{k}}$ es

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

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 13 (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 $\frac{Y}{c}$.

Teorema 14. 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

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

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 $\frac{Z}{1}$. \square

Programas usados en el documento

Programa 1 (El método congruencial mixto).

```
## Iniciaremos con un conjunto de parámetros para a, c y m
## que produzcan una secuencia periódica pequeña

a<-137; c<-0; m<-256;
x <- c(rep(0,each=512))
for (j in 1:256)
  x[j+1] <- (a*x[j] +c) %% m;
u=x/m;

a=314159269; c=453806245; m=2^31-1;
x = c(rep(0,each=257))
x[1]=87;
for (j in 1:256)
  x[j+1] = (a*x[j]+c)%%m;
u=x/m;

jpeg(filename="mixto.jpeg", width = 22, height = 22, res= 72, units = "cm")
plot(u[2:255],u[1:254])
dev.off()
```

Programa 2 (El método del rechazo para una normal estándar).

```
##### Normal #####
### Metodo del rechazo ###
# X:valores de la VA
# P:funcion de densidad de X
# Q:funcion de densidad auxiliar para metodo del rechazo

rechazo <- function(){

y1 <- rexp(1); y2 <- rexp(1)
while ( y2-0.5*(y1-1)^2 <= 0 ){
  y1 <- rexp(1); y2 <- rexp(1);
}
y <- y2-0.5*(y1-1)^2; u <- runif(1)
if (u<0.5)
  z = y1;
if (u>0.5)
  z = -y1;
z
}

## Probando el método del rechazo para 50 000
prueba_rechazo <- function(k){
x<-1:k
for(i in 1:k)
  x[i] <- rechazo()
jpeg(filename="rechazo.jpeg", width = 30, height = 18, res= 72, units = "cm")
hist(x, probability = TRUE)
rug(x)
lines(density(x))
dev.off()
}
prueba_rechazo(50000)

## Probando el generador de R para 50 000
jpeg(filename="normal.jpeg", width = 30, height = 19, res= 72, units = "cm")
fnorm <- rnorm(50000,0,1)
hist(fnorm, probability = TRUE)
rug(fnorm)
lines(density(fnorm))
#dev.off()
```

Programa 3 (El método polar).

```
##### Método polar #####
## El programa genera X,Y v.a. normales unitarias independientes

polar <- function(){
  u1 <- runif(1); u2 <- runif(1)
  v1 <- 2*u1; v2 <- 2*u2; s = v1^2 +v2^2

  while (s>1) {
    v1 <- runif(1,-1,1); v2 <- runif(1,-1,1); s = v1^2 +v2^2
  }

  r <- (-2.0*log(s)/s)^(0.5)
  x <- r*v1; y <- r*v2
  return(c(x,y))
}

prueba <- function(k){
  sx<-1:k; sy<-1:k
  for(i in 1:k){
    v <- polar()
    sx[i]<-v[1]; sy[i]<-v[2]
  }

  ## histograma en X
  jpeg(filename="polar_sx.jpeg", width = 30, height = 18, res= 72, units = "cm")
  hist(sx, probability = TRUE)
  rug(sx)
  lines(density(sx),col='blue')
  #lines(density(rnorm(50000,0,1)),col= 'red')
  dev.off()

  ## histograma en Y
  jpeg(filename="polar_sy.jpeg", width = 30, height = 18, res= 72, units = "cm")
  hist(sy, probability = TRUE)
  rug(sy)
  lines(density(sy),col='blue')
  #lines(density(rnorm(50000,0,1)),col= 'red')
  dev.off()
}

prueba(50000)
```

Programa 4 (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 5 (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 6 (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)
```


Bibliografía

- [1] Peter Dalgaard. *Introductory statistics with R*. Statistics and Computing. Springer-Verlag, New York, 2002.
- [2] Giuseppe Dodge, Yadolah; Melfi. *Premiers pas en simulation*. Springer, Paris, first edition, 2008.
- [3] Jacques Franchi. *Processus aléatoires à temps discret*. Ellipses Édition Marketing S.A., Paris, first edition, 2013.
- [4] James E. Gentle. *Random number generation and Monte Carlo methods*. Statistics and Computing. Springer, New York, second edition, 2003.
- [5] Sheldon M. Ross. *Introductory Statistics*. Elsevier/Academic Press, Amsterdam, 2010. 3rd edition.
- [6] Sheldon M. Ross. *Simulation*. Elsevier/Academic Press, Amsterdam, 2013. Fifth edition.