

4

SIMULACIÓN DE VARIABLES ALEATORIAS CON DISTRIBUCIÓN NORMAL

La distribución normal ha sido considerada clave en diversos procesos de inferencia estadística, en los que se requiere que se cumpla el supuesto de normalidad; esta distribución ha permitido también modelar gran cantidad de fenómenos que ocurren con cierta regularidad en la naturaleza, de allí que haya sido estudiada ampliamente tanto en su forma analítica, como a través de simulaciones; esta distribución teórica de probabilidad se ha constituido en una función límite o de convergencia, cuando se aplica el teorema central del límite (Mayorga, 2003).

El teorema central del límite establece que si X_1, X_2, X_3, \dots , es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media finita μ y varianza finita σ^2 , entonces

$$\lim_{n \rightarrow \infty} P\left(\frac{(X_1 + X_2 + \dots + X_n) - n\mu}{\sigma \sqrt{n}} \leq x\right) = \Phi(x) \quad \text{para todo } x \in R$$

Donde $\Phi(x)$ corresponde a la distribución normal estándar $N(0,1)$ (Ross, 1999). Ahora, si X_1, X_2, X_n, \dots , es una muestra aleatoria de una población con media finita μ y varianza finita σ^2 , entonces la sucesión de variables aleatorias

$$Z_n = \frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1)$$

converge a una distribución normal estándar (Mayorga, 2003).

Debido a que en muchos procesos de simulación se requiere simular valores de variables aleatorias con distribución normal con los parámetros fijos: media y una varianza finita

dadas, y puesto que dichos valores se pueden obtener por medio de los valores simulados a través de una variable aleatoria con distribución normal estándar, en este capítulo se presentan diversas formas de simular valores de variables aleatorias con distribución normal estándar utilizando los métodos propuestos por Papoulis (1991), Ríos (2000), Blanco (2004), Burbano (2010) y Marsaglia (1963); además, se presentan ejemplos de aplicación.

4.1 Distribución normal

De acuerdo con Bickel (1977), una variable aleatoria X sigue una distribución normal de parámetros μ y σ donde μ es un número real, y σ es un número real positivo si su función de densidad de probabilidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in R \quad (4.1)$$

De hecho, la función dada en (4.1) satisface las dos condiciones siguientes:

1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$

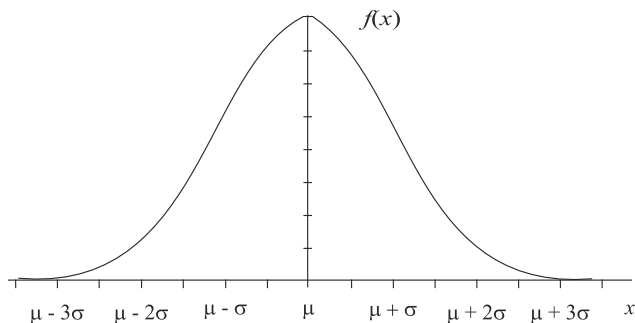


Figura 4.1. Curva normal o campana de Gauss

Fuente: Los autores, con el software libre R.

El valor esperado y la varianza son:

$$E(X) = \mu ; \text{Var}(X) = \sigma^2$$

La representación gráfica de la expresión (4.1) corresponde a una curva denominada campana de Gauss; esta se indica en la Figura 4.1.

Por ejemplo, si se tienen los parámetros $\mu = 600$ y $\sigma = 20$, la función de densidad de la distribución normal es:

$$f(x) = \frac{1}{20\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-600}{20}\right)^2\right], \quad x \in R$$

Como caso particular, si $\mu = 0$ y $\sigma = 1$, se tiene la densidad de la distribución normal estándar, definida por:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right], \quad x \in R$$

La anterior función de densidad de la distribución normal estándar para una variable aleatoria Z se suele escribir de la siguiente forma:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (4.2)$$

La expresión (4.2) se obtiene de la expresión (4.1) al realizar la sustitución siguiente:

$$z = \frac{x - \mu}{\sigma}$$

La función de distribución de la variable aleatoria Z con distribución normal estándar, que se denota con $Z \rightarrow N(0,1)$ es:

$$F(z) = P(Z \leq z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right] dx$$

4.1.1 Método de Box-Muller para simular una normal estándar

De acuerdo con Papoulis (1991), en concordancia con lo expuesto en la sección 2.7.3, para utilizar el método de Box-Muller se necesita considerar dos variables aleatorias independientes U_1, U_2 , cada una con distribución uniforme en el intervalo $(0,1)$, las cuales hacen posible la generación de los valores u_1, u_2 , pertenecientes a las variables ya mencionadas.

Para generar los valores z_1, z_2 , correspondientes a las variables aleatorias independientes Z_1, Z_2 , cada una con distribución normal estándar, se utilizan las expresiones 4.3 y 4.4.

$$z_1 = \sqrt{-2\ln(u_1)} \cos(2\pi u_2) \quad (4.3)$$

$$z_2 = \sqrt{-2\ln(u_1)} \operatorname{sen}(2\pi u_2) \quad (4.4)$$

A fin de proporcionar un ejemplo, se han generado 20 números aleatorios provenientes de una distribución uniforme en el intervalo $(0,1)$, los que se indican en las dos primeras columnas de la Tabla 4.1; a estos se les aplicó la prueba K-S de Kolmogorov-Smirnov para verificar si provienen de una distribución uniforme, y se obtuvo un P-valor de

0.52. Lo anterior prueba que, efectivamente, los 20 números aleatorios se ajustan a una distribución uniforme en (0,1). Así mismo, con un nivel de confianza del 95%, se obtuvo el intervalo (0.320 , 0.595) que no rechaza la hipótesis H_0 de que la media poblacional de la distribución de donde provienen los 20 números aleatorios es $\frac{1}{2}$.

Usando una asignación aleatoria, la secuencia de números aleatorios fue dividida en dos grupos, cada uno con 10 elementos, para posibilitar la aplicación de las expresiones (4.3) y (4.4). Los resultados se presentan en la Tabla 4.1. Los dos primeros valores aleatorios se obtuvieron del siguiente modo:

$$z_1 = \sqrt{-2Ln(0.119811)} \cos(2\pi (0.206008)) = 1.503366$$

$$z_2 = \sqrt{-2Ln(0.119811)} \text{sen}(2\pi (0.206008)) = 1.408396$$

Al aplicar la prueba K-S a los 20 valores simulados de la Tabla 4.1 se obtiene un P-valor de 0.635, el cual garantiza que dichos valores se ajustan a una distribución normal estándar.

En la eventualidad de que se requiera simular los valores x de una variable aleatoria X con distribución normal de parámetros μ y σ^2 , se utiliza la expresión

$$x = \mu + z\sigma$$

donde los z son los valores simulados de una variable aleatoria Z con distribución normal estándar.

Tabla 4.1. Valores simulados por el método de Box-Muller

u_1	u_2	z_1	z_2
0,119811	0,206008	1,503366	1,408396
0,058894	0,773391	2,218829	0,860712
0,188019	0,06692	0,694112	1,691339
0,099558	0,527075	1,741293	1,25775
0,67587	0,661839	-0,397529	-0,790872
0,692652	0,935801	-0,302155	-0,801971
0,866791	0,315688	0,358106	-0,307083
0,351231	0,522724	-0,859316	1,1637
0,496423	0,76585	-1,18319	0,026586
0,701978	0,127073	0,249991	-0,80325

(Valores generados por los autores con el software libre R)

4.1.2 Método de las 12 uniformes para una normal estándar

Siguiendo las ideas expuestas en Ríos (2000), para aplicar el método de las 12 uniformes se necesitan n variables aleatorias independientes U_1, U_2, \dots, U_n , cada una de ellas con distribución uniforme en el intervalo $(0,1)$; con estas variables se obtienen los valores u_1, u_2, \dots, u_n . Para generar un valor z correspondiente a la variable aleatoria Z con distribución normal estándar se utiliza la siguiente expresión, como consecuencia de la aplicación del teorema central del límite:

$$z = \frac{\sum_{i=1}^n u_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

La anterior igualdad se cumple, ya que cada U_1 tiene valor esperado $\frac{1}{2}$, y varianza $\frac{1}{12}$; además, las U_i son variables aleatorias independientes tales que su suma

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

tiene valor esperado y varianza dados por

$$E(S_n) = E\left(\sum_{i=1}^n U_i\right) = \frac{n}{2}$$

$$Var(S_n) = Var\left(\sum_{i=1}^n U_i\right) = \frac{n}{12}$$

y la variable aleatoria siguiente tiene distribución normal estándar:

$$Z = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$$

Entre más grande sea n , por el comportamiento asintótico de la expresión anterior, mayor es la probabilidad de que z corresponda a un valor proveniente de la variable Z con distribución normal estándar. Es razonable tomar $n = 12$ para que el valor z obtenido se considere que corresponde a una distribución normal estándar, admitiendo un error considerablemente pequeño. Sustituyendo en la expresión anterior, se tiene que:

$$z = \frac{\sum_{i=1}^{12} u_i - \frac{12}{2}}{\sqrt{\frac{12}{12}}} = \sum_{i=1}^{12} u_i - 6$$

Luego

$$z = \sum_{i=1}^{12} u_i - 6 \quad (4.5)$$

La expresión (4.5) representa el método de las 12 uniformes para generar valores de una variable Z con distribución normal estándar.

A continuación se simulan 10 valores de una distribución normal por el método de las 12 uniformes. Para esto se seleccionaron 10 muestras aleatorias, cada una de tamaño 12, provenientes del grupo de los 20 números aleatorios con distribución uniforme en $(0,1)$ presentados en la Tabla 4.1.

La primera muestra resultó conformada por : 0.119811, 0.058894, 0.099558, 0.675870, 0.866791, 0.496423, 0.701978, 0.206008, 0.773391, 0.026396.

$$z = \sum_{i=1}^{12} u_i - 6 = 3.223130 - 6 = -2.77687$$

Procediendo de manera similar para las 9 muestras restantes se obtienen los valores simulados que aparecen en la Tabla 4.2.

Tabla 4.2. Valores simulados por el método de 12 uniformes

Z	Z
-2,77687	-1,14753
-0,742584	0,348042
-0,629593	0,514155
1,623618	0,816556
-0,837836	0,659167

(Valores generados por los autores con el software libre R)

Mediante la prueba K-S aplicada a los 10 valores simulados de la Tabla 4.2. se obtiene un P-valor de 0.942 que permite concluir que dichos valores se ajustan a una distribución normal estándar.

4.1.3 Método del rechazo para simular una normal estándar

Con fundamento en lo expuesto en la sección 2.7.2 del capítulo 2, y en concordancia con Blanco (2004), de forma general, el método del rechazo exige, para generar valores x de

una variable aleatoria X con función de densidad de probabilidad f , que se genere una variable aleatoria Y con una función de densidad g . Posteriormente, se genera un número aleatorio, que llamaremos “ RND ”, perteneciente al intervalo $(0,1)$ y se acepta este valor generado con una probabilidad proporcional al cociente

$$\frac{f(Y)}{g(Y)}$$

Lo anterior permite desarrollar el siguiente algoritmo:

1. Se genera con densidad g
2. Se genera un número aleatorio RND
3. Si $RND \leq \frac{f(Y)}{cg(Y)}$, entonces hacer $X = Y$; en caso contrario, regresar a 1.

La variable aleatoria X tiene función de densidad f . Recuérdese que si X corresponde a una variable aleatoria con distribución normal estándar, entonces su función de densidad de probabilidad es:

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

$x \in R$, para generar los valores de X se observa en primer lugar que la variable aleatoria $W = |X|$ tiene la siguiente función de densidad:

$$h(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

$x \in R$.

Enseguida, se aplican los pasos que se indican a continuación: se genera primero una variable aleatoria W , se obtienen luego los valores de la variable X al hacer que sea igualmente probable que W sea igual a X o a $-X$.

Para generar W se usa el método del rechazo con $g(x) = \exp(-x)$ para $x > 0$; se observa que la función $\frac{h(x)}{g(x)}$ toma su valor máximo en $x = 1$.

Así,

$$c = \frac{2}{\sqrt{2\pi}} \exp\left(\frac{1}{2}\right)$$

Por consiguiente,

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

Así, los pasos del algoritmo para el método del rechazo quedan:

1. Generar Y con densidad exponencial de parámetro igual a 1
2. Generar un número aleatorio RND
3. Si $RND \leq \exp\left(-\frac{1}{2}(Y-1)^2\right)$, entonces hacer $X = Y$; en caso contrario regresar a 1

El paso 3 implica que si $-\ln(RND) \geq \frac{1}{2}(Y-1)^2$, entonces hacer $X = Y$. Puesto que $\ln(RND)$ sigue una distribución exponencial de parámetro igual a 1, el algoritmo se puede reescribir así:

1. Generar Y_1 con densidad exponencial de parámetro igual a 1
2. Generar Y_2 con densidad exponencial de parámetro igual a 1
3. Si $Y_2 \geq \frac{1}{2}(Y_1-1)^2$, entonces hacer $X = Y_1$; en caso contrario, regresar a 1

En Ross (1999) se demuestra que un algoritmo equivalente al anterior está conformado por los siguientes pasos:

1. Generar Y_1 con densidad exponencial de parámetro igual a 1
2. Generar Y_2 con densidad exponencial de parámetro igual a 1
3. Si $Y_2 \geq \frac{1}{2}(Y_1-1)^2$, entonces hacer $Y = Y_2 - \frac{1}{2}(Y_1-1)^2$ e ir al paso 4.
En caso contrario, regresar a 1
4. Generar un número aleatorio rnd y hacer:

$$z = \begin{cases} Y_1 & \text{si } rnd \leq 0.5 \\ -Y_1 & \text{si } rnd > 0.5 \end{cases}$$

Finalmente, el algoritmo anterior se puede escribir de la siguiente manera (Blanco, 2004):

1. Entrada: número máx. de simulaciones
Salida: Valores z correspondiente a la variable aleatoria Z
2. Para i desde 1 hasta máx. haga:

Inicialización:

Genere una variable aleatoria Y con distribución exponencial de parámetro $\lambda = 1$

Genere un número aleatorio RND

Verificación:

Mientras $RND > \exp(-\frac{1}{2}(Y-1)^2)$ haga:

Genere una variable aleatoria Y con distribución exponencial de parámetro $\lambda = 1$

Genere un número aleatorio RND

Salida: (Para escribir Y o $-Y$ con igual probabilidad)

Genere un número aleatorio rnd

Si $rnd > 0.5$ entonces escriba $-Y$ como valor de z
en caso contrario escriba Y como valor de z

Como ejemplo, a continuación se generan 17 valores z de una variable aleatoria Z , con distribución normal estándar, por el método del rechazo.

Para ello, de antemano se generaron 30 valores u provenientes de una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$, los cuales aparecen en la primera columna de la Tabla 4.3. Para estos, al aplicar la prueba K-S se obtuvo un P-valor de 0.237, el cual no rechaza la hipótesis H_0 de que dichos valores se ajustan a una distribución uniforme; así mismo, con un nivel de confianza del 95% se obtuvo en intervalo $(0.327, 0.572)$, que indica que la media de la distribución de donde provienen los datos es $\frac{1}{2}$.

En concordancia con la observación dada en la sección 3.2.2 para generar los valores y de una variable aleatoria Y con distribución exponencial de parámetro $\lambda = 1$, se puede usar la expresión

$$y = -\ln(u)$$

Los valores resultantes se indican en la segunda columna de la Tabla 4.3. Luego, se generan los números aleatorios "RND" que se presentan en la tercera columna de la tabla citada. Posteriormente se obtienen los resultados de la expresión

$$\exp(-\frac{1}{2}(Y-1)^2)$$

Estos aparecen en la cuarta columna de la mencionada tabla.

Cada valor de la tercera columna se compara con su correspondiente de la cuarta columna, y se toma la decisión de hacer $X = Y$ o de ir al paso 1. (Cada vez que se va al paso 1 no aparece valor en la sexta columna). Finalmente, con la ayuda del número aleatorio rnd , que se muestra en la quinta columna, se decide asignar el signo positivo

o negativo al valor de Y que aparece en la sexta columna de la tabla 4.3, el cual corresponde al valor z de la variable normal estándar que se está simulando.

Tabla 4.3. Valores simulados por el método del rechazo para una normal estándar

U	y	RND	$exp()$	rnd	Z
0,89368	0,112407	0,343307	0,674415	0,466829	0,112407
0,655186	0,422836	0,514829	0,846573	0,743976	-0,422836
0,006037	5,109919	0,260699	0,000215	0,539072	.
0,066359	2,712678	0,902257	0,230701	0,668639	.
0,039488	3,231759	0,23169	0,082879	0,930804	.
0,003585	5,630975	0,408919	0,000022	0,127935	.
0,219028	1,518557	0,621009	0,874196	0,748528	-1,51856
0,192394	1,64821	0,936747	0,810513	0,191731	.
0,951205	0,050025	0,010805	0,636847	0,045616	0,050025
0,674497	0,393788	0,354148	0,832147	0,383211	0,393788
0,557116	0,584982	0,71184	0,917484	0,304072	0,584982
0,572342	0,558019	0,828481	0,906945	0,40495	0,558019
0,098285	2,319882	0,61449	0,418514	0,898674	.
0,214664	1,53868	0,839025	0,864946	0,041893	1,53868
0,929609	0,072992	0,839595	0,650722	0,733871	.
0,82472	0,192711	0,974022	0,721908	0,162108	.
0,827113	0,189814	0,751837	0,720219	0,119993	.
0,614514	0,486923	0,40375	0,876671	0,668872	-0,486923
0,339462	1,080395	0,12773	0,996774	0,03097	1,080395
0,935187	0,067009	0,623687	0,647112	0,733664	-0,067009
0,235188	1,447371	0,833654	0,904774	0,268076	1,447371
0,827433	0,189427	0,495884	0,719992	0,255219	0,189427
0,267174	1,319854	0,975363	0,950133	0,144375	.
0,095979	2,343629	0,847684	0,405486	0,227685	.
0,138656	1,975758	0,717246	0,621231	0,509968	.
0,375488	0,979528	0,10467	0,99979	0,88466	-0,979528
0,247553	1,396131	0,038181	0,924539	0,088039	1,396131
0,253847	1,371024	0,282268	0,933486	0,784641	-1,37102
0,906985	0,097629	0,961999	0,665553	0,893541	.
0,530944	0,633099	0,011118	0,934907	0,646766	-0,633099

(Valores generados por los autores con el software libre R)

A los 17 valores simulados se les aplicó la prueba K-S, obteniéndose un P-valor de 0.999, que indica que dichos valores se ajustan a una distribución normal estándar.

4.1.4 Método de la transformada inversa usando la Distribución Lambda Generalizada (DLG)

En Burbano (2010) se indica una manera alternativa de generar valores z de la variable normal estándar Z por medio de la aplicación del método de la transformada inversa sobre la función inversa de la Distribución Lambda Generalizada citada por Karian y Dudewicz (2000). Este método se describe aquí de manera breve, y se usa para generar valores de una variable con distribución normal estándar.

La Distribución Lambda Generalizada se define por la siguiente expresión (Karian y Dudewicz, 2000):

$$F^{-1}(y) = F^{-1}(y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1 - y^{\lambda_4})}{\lambda_2}$$

con $\lambda_2 \neq 0$, $0 \leq y \leq 1$, donde λ_1 es el parámetro de localización, λ_2 es el parámetro de escala, λ_3 determina el sesgo (coeficiente de asimetría) y λ_4 determina la curtosis.

La distribución normal estándar se puede obtener asignando a los anteriores parámetros los siguientes valores específicos (Karian y Dudewicz, 2000):

$$\lambda_1 = 0, \lambda_2 = 0.1975, \lambda_3 = 0.1349, \lambda_4 = 0.1349$$

Con su reemplazo resulta:

$$F^{-1}(y) = \frac{y^{0.1349} - (1 - y)^{0.1349}}{0.1975} = x$$

Los valores z de la variable aleatoria Z con distribución normal estándar se obtienen generando valores u , correspondientes a una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$ que reemplazan a los valores “ y ” en la anterior expresión; esto es posible como una aplicación directa del método de la transformada inversa, es decir,

$$z = F^{-1}(u) = \frac{u^{0.1349} - (1 - u)^{0.1349}}{0.1975} \quad (4.6)$$

Para ejemplificar este método se requiere simular 20 valores z de la variable aleatoria Z con distribución normal estándar; para ello se usarán los 20 primeros valores u de la Tabla 4.3, los cuales provienen de una variable aleatoria U con distribución uniforme

en (0,1). Al reemplazar directamente en la expresión (4.6) se obtienen los 20 valores requeridos. Los resultados se presentan en la Tabla 4.4.

Al aplicar la prueba K-S a los 20 valores simulados se obtuvo un P-valor de 0.807, que indica que se ajustan a una distribución normal estándar.

Tabla 4.4. Valores simulados usando la Distribución Lambda Generalizada

u	z	U	z
0,89368	1,244925	0,557116	0,142622
0,655186	0,396718	0,572342	0,181031
0,006037	-2,51782	0,098285	-1,29041
0,066359	-1,50499	0,214664	-0,786682
0,039488	-1,76172	0,929609	1,474006
0,003585	-2,692	0,82472	0,930113
0,219028	-0,771816	0,827113	0,939455
0,192394	-0,865547	0,614514	0,289082
0,951205	1,66029	0,339462	-0,411216
0,674497	0,449463	0,935187	1,517261

(Valores generados por los autores con el software libre R)

Tabla 4.4b. Valores simulados aplicando el método de Marsaglia

U_1	U_2	V_1	V_2	T	W	X	Y
0,559355	0,761206	0,11871	0,522412	1,282244	N	.	.
0,982616	0,803134	0,965232	0,606268	3,143	N	.	.
0,325262	0,702858	-0,34948	0,405716	0,11248	6,23306	-2,1783	2,528851
0,535503	0,047913	0,071006	-0,90417	-1,66634	N	.	.
0,507678	0,394261	0,015356	-0,21148	-0,39224	N	.	.
0,281995	0,22862	-0,43601	-0,54276	-1,95754	N	.	.
0,347206	0,726215	-0,30559	0,45243	0,293684	2,8886	-0,88272	1,306889
0,921085	0,882465	0,84217	0,76493	3,2142	N	.	.
0,408148	0,182222	-0,1837	-0,63556	-1,63852	N	.	.
0,095739	0,27969	-0,80852	-0,44062	-2,49828	N	.	.
0,333891	0,695545	-0,33222	0,39109	0,117744	6,02803	-2,00262	2,357503
0,620771	0,902572	0,241542	0,805144	2,093372	N	.	.

(Pasa Tabla 4.4b)

(Viene Tabla 4.4b)

U_1	U_2	V_1	V_2	T	W	X	Y
0,641131	0,122813	0,282262	-0,75437	-0,94422	N	.	.
0,752135	0,507697	0,50427	0,015394	1,039328	N	.	.
0,204952	0,117283	-0,5901	-0,76543	-2,71106	N	.	.
0,143259	0,924725	-0,71348	0,84945	0,271936	3,0947	-2,20801	2,628793
0,445881	0,517283	-0,10824	0,034566	-0,14734	N	.	.
0,369064	0,271406	-0,26187	-0,45719	-1,43812	N	.	.
0,48528	0,644636	-0,02944	0,289272	0,519664	1,5872	-0,04673	0,459134
0,099268	0,643786	-0,80146	0,287572	-1,02778	N	.	.
0,725651	0,852829	0,451302	0,705658	2,31392	N	.	.
0,037251	0,06515	-0,9255	-0,8697	-3,5904	N	.	.
0,174697	0,406798	-0,65061	-0,1864	-1,67402	N	.	.
0,288352	0,566701	-0,4233	0,133402	-0,57979	N	.	.
0,66887	0,84329	0,33774	0,68658	2,04864	N	.	.

(Valores generados por los autores con el software libre R)

4.1.5 Aplicación del método de Marsaglia

A continuación se simulan algunos valores x , y de una variable aleatoria Z con distribución normal estándar, aplicando el algoritmo de Marsaglia descrito en la parte final de la sección 2.7.3. Para ello se generaron U_1 y U_2 con distribución uniforme en el intervalo $(0,1)$, que aparecen en las dos primeras columnas de la Tabla 4.4b; se calcularon V_1 , V_2 , T , cuyos resultados aparecen en las columnas tercera, cuarta y quinta, respectivamente. Sin embargo, el valor de W solo se pudo calcular para aquellos valores de T no mayores que 1 y que fueran positivos; en los casos marcados con una N no fue posible realizar el cálculo de W , y, por consiguiente, no se generaron los valores de la distribución normal estándar; finalmente, dichos valores se generaron para cinco parejas de valores de la distribución uniforme U_1 y U_2 , lográndose simular 10 valores de la distribución normal estándar.

4.2 Ejemplos de aplicación de la distribución normal

En esta sección se presentan algunos ejemplos de aplicación sobre la simulación de valores de una distribución normal con parámetros conocidos. Para esto, inicialmente se va a generar una cantidad suficiente de números pseudoaleatorios que se someterán a las pruebas estadísticas descritas en la sección 1.4, a fin de verificar si son números aleatorios; luego, se simularán valores aleatorios correspondientes a una distribución

normal estándar, y con ellos se generarán valores aleatorios de una distribución normal con parámetros específicos asociados a una situación hipotética.

4.2.1 Aplicación en una institución financiera

En una institución financiera se requiere simular los ingresos mensuales de los próximos 25 potenciales clientes que lleguen a ella a solicitar un crédito de libre inversión. De acuerdo con los perfiles presentados por los últimos clientes que han solicitado ese tipo de crédito, se maneja el supuesto de que los ingresos mensuales de los potenciales clientes pueden provenir de una población normal con media $\mu = 620000$ pesos mensuales y desviación estándar $\sigma = 15000$ pesos por mes.

4.2.1.1 Generación de números aleatorios. Se inicia generando una tabla con 50 números pseudoaleatorios u con distribución uniforme $U(0,1)$, los cuales se presentan en la Tabla 4.5. A dichos números se les aplica la prueba de la media, de la varianza y la de independencia, con el fin de comprobar que dichos números corresponden efectivamente a números aleatorios y poderlos utilizar para generar los valores x de las variables aleatorias X de interés, que se describen y se ejemplifican en las siguientes subsecciones.

El signo más (+) y el signo menos (-) de la Tabla 4.5 se usarán en la prueba de independencia. Se generan 50 números, con el fin de tener la posibilidad de seleccionar de ese grupo los 25 que se necesitan para desarrollar el proceso de simulación.

El intervalo de confianza del 95% se denota con (L_r, L_s) , donde:

$$\alpha = 0.05$$

$$L_i = \frac{\chi_{\frac{\alpha}{2}}^2}{12(n-1)} \quad L_s = \frac{\chi_{1-\frac{\alpha}{2}}^2}{12(n-1)}$$

$$S^2 = \frac{\sum_{i=1}^{50} (r_i - \bar{X})^2}{49} = 0.075$$

Tabla 4.5 Números pseudoaleatorios generados mediante una distribución uniforme en (0,1)

N.º	Pseudo-aleatorio	Signo	N.º	Pseudo-aleatorio	Signo	N.º	Pseudo-aleatorio	Signo
1	0,559355		18	0,369064	-	35	0,27969	+
2	0,982616	+	19	0,48528	+	36	0,695545	+
3	0,325262	-	20	0,099268	-	37	0,902572	+
4	0,535503	+	21	0,725651	+	38	0,122813	-
5	0,507678	-	22	0,037251	-	39	0,507697	+
6	0,281995	-	23	0,174697	+	40	0,117283	-
7	0,347206	+	24	0,288352	+	41	0,924725	+
8	0,921085	+	25	0,66887	+	42	0,517283	-
9	0,408148	-	26	0,761206	+	43	0,271406	-
10	0,095739	-	27	0,803134	+	44	0,644636	+
11	0,333891	+	28	0,702858	-	45	0,643786	-
12	0,620771	+	29	0,047913	-	46	0,852829	+
13	0,641131	+	30	0,394261	+	47	0,06515	-
14	0,752135	+	31	0,22862	-	48	0,406798	+
15	0,204952	-	32	0,726215	+	49	0,566701	+
16	0,143259	-	33	0,882465	+	50	0,84329	+
17	0,445881	+	34	0,182222	-			

(Valores generados por los autores con el software libre R)

A los números pseudoaleatorios presentados en la Tabla 4.5 se les aplicó la prueba K-S, de Kolmogorov-Smirnov, para verificar si se ajustan a una distribución uniforme en el intervalo (0,1); se obtuvo un P-valor de 0.946, lo cual garantiza que los mencionados números pseudoaleatorios efectivamente se ajustan a una distribución uniforme en el intervalo (0,1) con media $\frac{1}{2}$. Además, se determinó un intervalo de confianza del 95% para la media, que resultó ser (0.4030, 0.5588), lo que prueba la hipótesis de que los números pseudoaleatorios citados provienen de una distribución uniforme con media 0.5. En este caso, también se podrían contrastar las hipótesis $H_0: \mu = 0.5$, $H_1: \mu \neq 0.5$.

Para la prueba de varianza se calcularon la media y la varianza de los números pseudoaleatorios generados, obteniéndose:

$$\bar{X} = 0.480922$$

Para los 50 números pseudoaleatorios quedan:

$$L_i = \frac{\chi_{0.025,49}^2}{12(49)} = \frac{31.5549}{588} = 0.0536$$

$$L_i = \frac{\chi_{0.975,49}^2}{12(49)} = \frac{70.2224}{588} = 0.1194$$

El intervalo (0.0536 , 0.1194) contiene a la varianza poblacional y también al valor $S^2 = 0.075$, lo cual indica que los 50 números pseudoaleatorios sí cumplen con la prueba de varianza.

Para la prueba de independencia, el número de corridas observadas h es igual a 31 y el tamaño de la muestra n es de 50; el valor esperado y la varianza se obtienen con las siguientes expresiones:

$$E(h) = \frac{2n-1}{3} \quad \text{Var}(h) = \frac{16n-29}{90}$$

$$E(h) = \frac{2(50)-1}{3} = \frac{99}{3} = 33$$

$$\text{Var}(h) = \frac{16(50)-29}{90} = \frac{771}{90} = 8.5666$$

Calculando el estadístico de prueba, se tiene:

$$Z = \frac{h - E(h)}{\sqrt{\text{Var}(h)}} = \frac{31 - 33}{\sqrt{8.5666}} = \frac{-2}{2.9269} = -0.6833$$

Si se cumple la condición $Z < Z_{\frac{\alpha}{2}}$ o $Z > Z_{1-\frac{\alpha}{2}}$ para un nivel de significancia igual al 0.05 y de acuerdo con la Figura 4.2, entonces se rechaza la hipótesis H_0 de que los números pseudoaleatorios son independientes, en caso contrario no se rechaza la H_0 de independencia.

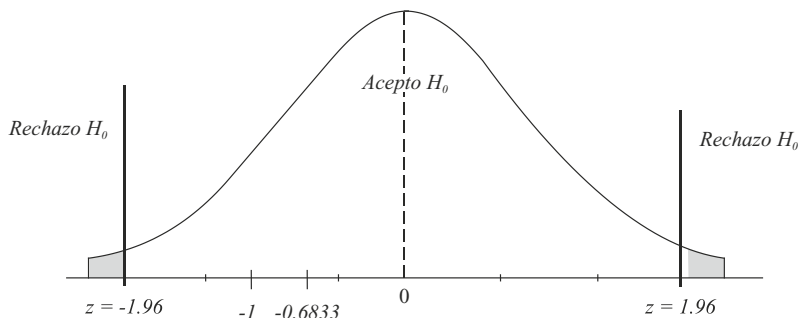


Figura 4.2. Regiones de aceptación y rechazo de H_0

(Esta gráfica de la distribución normal estándar fue elaborada por los autores con el software libre R).

En la Figura 4.2. se observa que el valor de la estadística de prueba es de $Z = -0.6833$, el cual cae en la región de aceptación (no rechazo) de H_0 , luego se concluye que los 50 números pseudoaleatorios son independientes. Del anterior análisis se concluye que los 50 números generados de la distribución uniforme son aleatorios y se pueden utilizar en procesos de simulación. Es importante aclarar que algunos números aleatorios de la Tabla 4.5 se han utilizado en los ejemplos desarrollados en el capítulo 3.

Aprovechando las bondades del computador, es posible generar miles e incluso millones de números aleatorios para realizar los procesos de simulación; además, se puede incluir una cantidad mayor de dígitos en cada número aleatorio, a fin de lograr una mejor precisión en los valores simulados correspondientes a las variables aleatorias de interés.

4.2.1.2 Simulación de valores de una variable normal estándar. Para realizar la simulación de los valores de una variable normal estándar se puede aprovechar cualquiera de los métodos ejemplificados en las subsecciones 4.1.1, 4.1.2, 4.1.3, 4.1.4 y 4.1.5. Por su versatilidad, aquí se utilizó la expresión 4.6 y se generaron los 50 valores que se indican en la Tabla 4.6.

Tabla 4.6. Valores simulados z que corresponden a una distribución normal estándar

N.º	u	x	N.º	u	x
1	0,559355	0,148256	26	0,761206	0,706543
2	0,982616	2,120229	27	0,803134	0,84928
3	0,325262	-0,450129	28	0,702858	0,529403
4	0,535503	0,08846	29	0,047913	-1,6692
5	0,507678	0,019105	30	0,394261	-0,266352
6	0,281995	-0,573543	31	0,22862	-0,739735
7	0,347206	-0,390267	32	0,726215	0,597957
8	0,921085	1,412776	33	0,882465	1,185467
9	0,408148	-0,230665	34	0,182222	-0,903458
10	0,095739	-1,3054	35	0,27969	-0,580362
11	0,333891	-0,426401	36	0,695545	0,508475
12	0,620771	0,305383	37	0,902572	1,295419
13	0,641131	0,359046	38	0,122813	-1,15889
14	0,752135	0,677615	39	0,507697	0,019153
15	0,204952	-0,820431	40	0,117283	-1,18676
16	0,143259	-1,063	41	0,924725	1,438269
17	0,445881	-0,13509	42	0,517283	0,043018

(Pasa Tabla 4.6)

(Viene Tabla 4.6)

N.º	u	x	N.º	u	x
18	0,369064	-0,332051	43	0,271406	-0,605099
19	0,48528	-0,036636	44	0,644636	0,368389
20	0,099268	-1,2847	45	0,643786	0,366119
21	0,725651	0,596266	46	0,852829	1,045753
22	0,037251	-1,78895	47	0,06515	-1,51456
23	0,174697	-0,932379	48	0,406798	-0,23412
24	0,288352	-0,554884	49	0,566701	0,166772
25	0,66887	0,433966	50	0,84329	1,004976

(Valores generados por los autores con el software libre R)

4.2.1.3 Simulación de ingresos. Para realizar la simulación solicitada en el problema de aplicación se pueden aprovechar los 25 primeros valores z , simulados para la variable normal estándar Z mediante la expresión (4.6), presentados en la Tabla 4.6. En estas circunstancias es conveniente utilizar la transformación que se indica a continuación:

$$z = \frac{x - \mu}{\sigma} \rightarrow x = z\sigma + \mu$$

Tabla 4.7. Ingresos simulados para 25 potenciales clientes

Cientes	z	Ingresos simulados
1	0,148256	622223,8
2	2,120229	651803,4
3	-0,450129	613248,1
4	0,08846	621326,9
5	0,019105	620286,6
6	-0,573543	611396,9
7	-0,390267	614146
8	1,412776	641191,6
9	-0,230665	616540
10	-1,3054	600419
11	-0,426401	613604
12	0,305383	624580,7
13	0,359046	625385,7
14	0,677615	630164,2

(Pasa Tabla 4.7)

(Viene Tabla 4.7)

Clientes	z	Ingresos simulados
15	-0,820431	607693,5
16	-1,063	604055
17	-0,13509	617973,7
18	-0,332051	615019,2
19	-0,036636	619450,5
20	-1,2847	600729,5
21	0,596266	628944
22	-1,78895	593165,8
23	-0,932379	606014,3
24	-0,554884	611676,7
25	0,433966	626509,5

(Valores generados por los autores con el software libre R)

Para ilustrar la forma como se obtienen los 24 ingresos simulados que se presentan en la Tabla 4.7, a continuación se indican los cálculos para simular los dos primeros ingresos utilizando los dos primeros números aleatorios z de la Tabla 4.6.

$$\text{Ingreso}_1 = 0,148255(15000) + 620000 = 622223,8$$

$$\text{Ingreso}_2 = 2,120237(15000) + 620000 = 651803,4$$

Ahora, si la institución financiera requiere simular los ingresos mensuales para los 500 potenciales clientes que se espera lleguen a solicitar el mismo crédito próximamente, se procede a simular 500 valores z utilizando la expresión 4.6 con su correspondiente algoritmo, y luego se pasa, de manera similar, a la generación de la Tabla 4.7.

4.2.2 Uso de la normal estándar para simular una variable con distribución Birnbaum-Saunders

4.2.2.1 La distribución Birnbaum-Saunders. Con base en lo expuesto en Birnbaum y Saunders (1969a, 1969b) y Saunders (2007), una variable aleatoria X tiene distribución Birnbaum-Saunders de parámetros $\alpha > 0$, $\beta > 0$, si su función de distribución está dada por:

$$F_x(x) = \Phi \left(\frac{1}{\alpha} \left[\left(\frac{x}{\beta} \right)^{\frac{1}{2}} - \left(\frac{\beta}{x} \right)^{\frac{1}{2}} \right] \right) = \Phi(h(x))$$

con

$$h(x) = z = \frac{1}{\alpha} \left[\left(\frac{x}{\beta} \right)^{\frac{1}{2}} - \left(\frac{\beta}{x} \right)^{\frac{1}{2}} \right] \quad (47)$$

La cual se denota con $X \rightarrow BS(\alpha, \beta)$; donde $\Phi(\cdot)$ representa la distribución normal estándar. En este caso, $z = h(x)$ corresponde al conjunto de los valores de una variable normal estándar Z . El p percentil de la distribución de la variable aleatoria $X \rightarrow BS(\alpha, \beta)$ es $x_p = F_X^{-1}(p)$ (Kandu, 2008, Díaz, 2005; Leiva, 2009; Barros, 2009); también este se puede obtener aplicando el método de la transformada inversa, descrito en la sección 2.7, subsección 2.7.1; de allí se deduce que

$$x_p = \beta \left[\frac{\alpha z_p}{2} + \frac{\sqrt{(\alpha z_p)^2 + 4}}{2} \right]^2$$

Donde z_p es el p percentil de la distribución de una variable normal estándar denotada con $Z \rightarrow N(0,1)$. La mencionada expresión se puede utilizar para generar valores de una variable aleatoria con distribución Birnbaum-Saunders, $X \rightarrow BS(\alpha, \beta)$.

4.2.2.2 Generación de un algoritmo de simulación. Se puede también utilizar un procedimiento similar al desarrollado por Chang y Tang (1994); para ello, de la expresión (4.7) se deduce que

$$Z^2 = \left(\frac{1}{\alpha} \right)^2 \left[\left(\frac{X}{\beta} \right) + \left(\frac{\beta}{X} \right) - 2 \right] \quad (4.8)$$

Resolviendo la ecuación 4.7 se obtiene

$$X^2 - \beta (2 + (\alpha Z)^2)X + \beta^2 = 0$$

Para $Z \neq 0$, las soluciones de la anterior ecuación cuadrática son:

$$X_1 = \beta \left[1 + \frac{(\alpha Z)^2}{2} + \sqrt{(\alpha Z)^2 + \frac{(\alpha Z)^4}{4}} \right]$$

$$X_2 = \beta \left[1 + \frac{(\alpha Z)^2}{2} - \sqrt{(\alpha Z)^2 + \frac{(\alpha Z)^4}{4}} \right]$$

Michael (1976) mostró que la probabilidad condicional de escoger cualquiera de las dos raíces $X_1 = x_1$ o $X_2 = x_2$, dado $Z = z$ es igual a 0.5. En consecuencia, para generar valores x de una variable $X \rightarrow BS(\alpha, \beta)$ es conveniente, primero, generar números aleatorios u con distribución uniforme en el intervalo $(0,1)$ que puedan utilizarse como juez (criterio) en la escogencia de una de las dos raíces; segundo, generar valores aleatorios z de una variable normal estándar $Z \rightarrow N(0,1)$ mediante algún procedimiento adecuado, y, tercero, aplicar el siguiente algoritmo:

1. Generar un número aleatorio u de una variable $U \rightarrow U(0,1)$
2. Generar un valor aleatorio z de una variable $Z \rightarrow N(0,1)$ mediante algún método de simulación
3. Para determinados valores de $\alpha > 0$ y $\beta > 0$, calcular el valor x de la variable aleatoria $X \rightarrow BS(\alpha, \beta)$ mediante el siguiente criterio:

Si $u < 0.5$, entonces

$$x = x_1 = \beta \left[1 + \frac{(\alpha z)^2}{2} + \sqrt{(\alpha z)^2 + \frac{(\alpha z)^4}{4}} \right] \quad (4.9)$$

Si $u > 0.5$, entonces

$$x = x_2 = \beta \left[1 + \frac{(\alpha z)^2}{2} - \sqrt{(\alpha z)^2 + \frac{(\alpha z)^4}{4}} \right] \quad (4.10)$$

Si $u = 0.5$, $z = 0$, entonces

$$x_{0.5} = \beta$$

Repetir los pasos anteriores n veces para generar los n valores que se requieren.

4.2.2.3 Un ejemplo numérico para un caso particular. Ahora se tomará como ejemplo una distribución Birnbaum-Saunders con los siguientes parámetros: $\alpha = 0.3$, $\beta = 1$, para simular 17 de sus valores.

Un algoritmo para simular los valores x de una variable $X \rightarrow BS(0.3,1)$, con base en las expresiones (4.9) y (4.10), incluye las siguientes decisiones:

Si $u < 0.5$, entonces

$$x = x_1 = \left[1 + \frac{(0.3z)^2}{2} + \sqrt{(0.3z)^2 + \frac{(0.3z)^4}{4}} \right]$$

Si $u > 0.5$, entonces

$$x = x_2 = \left[1 + \frac{(0.3z)^2}{2} - \sqrt{(0.3z)^2 + \frac{(0.3z)^4}{4}} \right]$$

Si $u = 0.5$, $z = 0$, entonces

$$x_{0.5} = \beta = 1$$

Enseguida se pueden aprovechar los 17 primeros números aleatorios u con distribución uniforme $U(0,1)$ de la Tabla 4.6 como criterio de decisión; luego, se tienen que generar 17 valores de una variable con distribución normal estándar; estos se pueden obtener utilizando cualquiera de las técnicas expuestas en la sección 4.1. Otra posibilidad consiste en utilizar los 17 primeros valores z de la Tabla 4.6 que ya pertenecen a una variable Z con distribución normal estándar. Al aplicar el algoritmo expuesto que incluye las expresiones (4.9) y (4.10) se obtienen los valores que aparecen en la tercera columna de la Tabla 4.8.

Tabla 4.8. Valores simulados de una variable con distribución $BS(0.3,1)$

u	z	$BS(0.3,1)$
0,559355	0,148255	0,956502
0,982616	2,120237	0,534829
0,325262	-0,450128	1,144464
0,535503	0,088461	0,973812
0,507678	0,019106	0,994285
0,281995	-0,573544	1,187502
0,347206	-0,390266	1,124134
0,921085	1,412778	0,656572
0,408148	-0,230666	1,071636
0,095739	-1,3054	1,47574
0,333891	-0,426401	1,136363
0,620771	0,305383	0,912486
0,641131	0,359046	0,897931
0,752135	0,677613	0,816331
0,204952	-0,820432	1,278276
0,143259	-1,063	1,373777
0,445881	-0,135089	1,041356

(Valores generados por los autores con el software libre R)

En la eventualidad de que se requiera simular 100 valores de una variable aleatoria con distribución Birnbaum-Saunders de parámetros 0.3 y 1, es decir, de $X \rightarrow BS(0.3,1)$, se simulan 10 valores z de una variable normal estándar y luego se utiliza el algoritmo que incluye las expresiones (4.9) y (4.10). De similar modo, se procede para un número mayor de simulaciones, 1.000, 5.000, 30.000, etc. En ese caso es conveniente escribir un programa para que se ejecute automáticamente en el computador.

Ejercicios capítulo 4

- 4.1 Simular 50 valores de una variable aleatoria con distribución normal estándar por el método de Box-Muller.
- 4.2 Simular 40 valores de una variable aleatoria con distribución normal estándar por el método de las 12 uniformes.
- 4.3 Simular 60 valores de una variable aleatoria con distribución normal estándar por el método del rechazo.
- 4.4 Simular 100 valores de una variable aleatoria con distribución normal estándar por el método de la DLG.
- 4.5 Simular 200 valores de una variable aleatoria con distribución normal estándar por el método de Marsaglia.
- 4.6 Simular 50 valores de una variable aleatoria con distribución normal de media 1.000.000 y desviación estándar de 50.000.
- 4.7 Simular 20 valores de una variable aleatoria con distribución Birnbaum-Saunders de parámetros
- 4.8 Las ventas en miles de pesos por día de las pequeñas empresas en la región R del departamento de Cundinamarca, Colombia, se distribuyen normalmente con media 20.000 y desviación estándar de 500. Simular 30 valores de las ventas para el próximo día, en 30 diferentes pequeñas empresas.