

Simulación con modelos aleatorios

Simulación con modelos aleatorios

Conocimiento estadístico-probabilístico y simulación

Víctor Miguel Ángel Burbano Pantoja
Margoth Adriana Valdivieso Miranda
Luis Alfonso Salcedo Plazas



Universidad Pedagógica y Tecnológica de Colombia
Tunja
2014

Simulación con modelos aleatorios: conocimiento estadístico-probabilístico y simulación / Víctor Miguel Ángel Burbano Pantoja; Margoth Adriana Valdivieso Miranda; Luis Alfonso Salcedo Plazas. – Tunja: Uptc, 2014.
148 p.: il.

ISBN 978-958-660-207-5 (rústico)

ISBN 978-958-660-208-2 (on line)

1. Simulación. – 2. Modelos aleatorios – I. Burbano Pantoja, Víctor Miguel Ángel. – II. Valdivieso Miranda, Margoth Adriana. – III. Salcedo Plazas, Luis Alfonso. – IV. Tit.

CDD 519.5/B946

Primera edición, 2014
100 ejemplares (impresos)

Simulación con modelos aleatorios: conocimiento estadístico-probabilístico y simulación

ISBN 978-958-660-207-5 (rústica)

ISBN 978-958-660-208-2 (on line)

Colección Facultad de Educación 70 años

© Víctor Miguel Ángel Burbano Pantoja
© Margoth Adriana Valdivieso Miranda
© Luis Alfonso Salcedo Plazas
© Universidad Pedagógica y Tecnológica de Colombia

Gustavo Orlando Álvarez Álvarez, Rector, Uptc

Comité Editorial

Celso Antonio Vargas Gómez, MSc
Hugo Alfonso Rojas Sarmiento, Ph.D.
Fred Manrique Abril, Ph.D.
Liliana Fernández Samacá, Ph.D.
Luz Eliana Márquez, MSc.
Fanor Casierra Posada, Ph.D.
Jovanny Arles Gómez Castaño, Ph.D.
Rigaud Sanabria Marín, Ph.D.
Pablo Enrique Pedraza Torres, Ph.D.

Coordinadora editorial: Yolanda Romero A.
Corrección de estilo: Luis Enrique Clavijo M.

Libro financiado por la Dirección de Investigaciones de la Uptc

Se permite la reproducción parcial o total con la autorización expresa de los titulares del derecho de autor.

Este libro es registrado en Depósito Legal, según lo establecido en la Ley 44 de 1993, el Decreto 460 del 16 de marzo de 1995, el Decreto 2150 de 1995 y el Decreto 358 de 2000.

Citación sugerida: Burbano, V.M.A. et. al. (2014). Simulación con modelos aleatorios: conocimiento estadístico-probabilístico y simulación. Tunja: Editorial Uptc, 148 p.

Impresión: Grupo Imprenta y Publicaciones
Rafael Humberto Parra Niño, Coordinador
UPTC - Avenida Central del Norte
Tels.: (0*8) 740 5626 - Exts. 2366 - 2367 - Fax 2408
imprenta.publicaciones@uptc.edu.co
Tunja - Boyacá – Colombia

Editorial Uptc
Edificio Administrativo – Piso 4
Av. Central del Norte
comite.editorial@uptc.edu.co
www.uptc.edu.co
(+57 8) 7425268

CONTENIDO

INTRODUCCIÓN	9
1. ELEMENTOS CONCEPTUALES SOBRE SIMULACIÓN	13
1.1 El concepto de simulación	13
1.2 Noción de lo aleatorio y de los números aleatorios	16
1.3 Generación de números aleatorios	17
1.3.1 <i>Método de los cuadrados medios</i>	17
1.3.2 <i>Método de las congruencias lineales</i>	19
1.4 Pruebas estadísticas para números pseudoaleatorios	20
1.4.1 <i>Prueba de la media</i>	21
1.4.2 <i>Prueba de la varianza</i>	25
1.4.3 <i>Prueba de independencia</i>	26
1.5 Planeación de experimentos de simulación en PC	28
1.5.1 <i>Formulación del problema</i>	28
1.5.2 <i>Recolección y procesamiento de datos</i>	29
1.5.3 <i>Modelo matemático</i>	30
1.5.4 <i>Estimación de parámetros</i>	31
1.5.5 <i>Evaluación del modelo y los parámetros</i>	32
1.5.6 <i>Formulación de un programa para la PC</i>	32
1.5.7 <i>Validación</i>	33
1.5.8 <i>Diseño de los experimentos de simulación</i>	33
1.5.9 <i>Análisis de los datos simulados</i>	34
1.6 Formas de utilizar el concepto de simulación	34
1.6.1 <i>Nivel intuitivo</i>	34
1.6.2 <i>Nivel formal</i>	35
1.6.3 <i>Conocimiento estadístico-probabilístico y simulación</i>	35
Ejercicios capítulo 1	36
2. CONOCIMIENTO PROBABILÍSTICO USUAL EN SIMULACIÓN	37
2.1 Elementos de probabilidad	37
2.1.1 <i>Experimento aleatorio</i>	37
2.1.2 <i>Sigma álgebra</i>	38
2.1.3 <i>Medida de probabilidad</i>	38
2.2 Variable aleatoria real	39

2.3	Función de probabilidad para variables aleatorias discretas	40
2.4	Función de densidad de probabilidad	41
2.5	Función de distribución	41
2.6	Esperanza matemática y varianza de una variable aleatoria	42
2.7	Métodos de simulación de una variable aleatoria	44
2.7.1	<i>Método de la transformada inversa</i>	45
2.7.2	<i>Método del rechazo</i>	45
2.7.3	<i>Método de Box Muller</i>	47
2.7.4	<i>Método de la composición</i>	51
2.7.5	<i>Método de la convolución</i>	52
2.8	Método de simulación Monte Carlo	52
2.8.1	<i>Concepto de muestra aleatoria</i>	53
2.8.2	<i>Ley fuerte de los grandes números</i>	53
2.8.3	<i>Un algoritmo para la simulación Monte Carlo</i>	53
2.8.4	<i>Aproximación de una integral en el intervalo (0,1)</i>	54
2.8.5	<i>Método de Monte Carlo para integrales en (a,b)</i>	56
2.8.6	<i>Aproximación de una integral en el intervalo (-2,2)</i>	56
	Ejercicios capítulo 2	58
3.	SIMULACIÓN CON VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS	59
3.1	Uso de modelos aleatorios discretos	59
3.1.1	<i>Modelo de Bernoulli</i>	59
3.1.2	<i>Modelo Binomial</i>	61
3.1.3	<i>Modelo de Poisson</i>	63
3.1.4	<i>Aplicación del método de la transformada inversa</i>	66
3.2	Simulación de variables aleatorias continuas	69
3.2.1	<i>Modelo uniforme</i>	69
3.2.2	<i>Modelo exponencial</i>	71
3.2.3	<i>Modelo continuo no usual</i>	74
3.2.4	<i>Aplicación del método de composición</i>	76
3.2.5	<i>Simulación con una distribución beta</i>	77
3.2.6	<i>Aplicación del método del rechazo</i>	80
	Ejercicios capítulo 3	82
4.	SIMULACIÓN DE VARIABLES CON DISTRIBUCIÓN NORMAL	83
4.1	Distribución normal	84
4.1.1	<i>Método de Box Muller para una normal estándar</i>	85
4.1.2	<i>Método de las 12 uniformes</i>	87
4.1.3	<i>Método del rechazo para una normal estándar</i>	88
4.1.4	<i>Método de la transformada inversa DLG normal est.</i>	93
4.1.5	<i>Aplicación del método de Marsaglia</i>	95
4.2	Ejemplos de aplicación de una distribución normal	95

4.2.1	<i>Aplicación en una institución financiera</i>	96
4.2.2	<i>Uso de una normal para simular una Birnbaum-S</i>	100
	Ejercicios capítulo 4	105
5.	SIMULACIÓN CON LA DISTRIBUCIÓN LAMBDA GENERALIZADA (DLG)	106
5.1	Caracterización de la DLG	106
5.1.1	<i>Definición de la DLG</i>	106
5.1.2	<i>Momentos de la DLG</i>	108
5.1.3	<i>Distribuciones simétricas provenientes de la DLG</i>	109
5.1.4	<i>Distribuciones asimétricas seleccionadas de la DLG</i>	113
5.1.5	<i>Distribuciones usuales provenientes de la DLG</i>	118
5.1.6	<i>Otras distribuciones seleccionadas de la DLG</i>	121
5.2	Simulación de variables aleatorias con DLG	123
5.2.1	<i>Algoritmo general para una DLG</i>	123
5.2.2	<i>Simulación con distribuciones simétricas</i>	124
5.2.3	<i>Simulación con distribuciones asimétricas</i>	127
5.2.4	<i>Simulación con distribución platicúrtica</i>	130
5.2.5	<i>Simulación con distribución leptocúrtica</i>	131
5.3	Ejemplos de aplicación de la simulación	135
5.3.1	<i>Variabilidad del espesor de un producto</i>	135
5.3.2	<i>Tiempos de servicio de unos clientes de supermercado</i>	136
5.3.3	<i>Utilidades generadas por unas pequeñas empresas</i>	137
5.3.4	<i>Tiempos entre llegadas de los clientes de un banco</i>	139
	Ejercicios capítulo 5	141
	REFERENCIAS	143

INTRODUCCIÓN

Los procesos de simulación se utilizan hoy, en distintos contextos y con variado nivel de complejidad, para explorar, experimentar y desarrollar proyectos de investigación en diversos campos del conocimiento humano. A pesar de que la palabra *simulación* generalmente se ha asociado con la aparición del computador, se intuye que el concepto de simulación emergió cuando la humanidad empezó a emular algunos fenómenos naturales que se suscitaban de manera aleatoria, o en el instante en el que pensó imitar el comportamiento de algunos animales para construir artefactos y máquinas que le permitieran vivir mejor.

Desde hace mucho tiempo, el ser humano ha inventado aparatos y desarrollado procedimientos que han simulado el desplazamiento y otros comportamientos de ciertas especies animales, para generar nuevas formas de transporte, para tratar de predecir la ocurrencia de fenómenos naturales, como la lluvia, la velocidad del viento o el movimiento de los cuerpos celestes, o para experimentar con procesos alternativos en la búsqueda de soluciones a los problemas que se generan en la sociedad.

De forma intuitiva, la simulación puede entenderse como emulación o imitación de situaciones reales con presencia del azar, de la incertidumbre, del riesgo, y al servicio de la experimentación o de la realización de cálculos aproximados, tanto de forma manual como electrónica. En la actualidad, el computador se ha constituido en una herramienta fundamental que posibilita en buena medida el desarrollo de gran variedad de procesos de investigación por medio de simulación. Los procesos de simulación por computador han crecido vertiginosamente, y se utilizan en diversos campos del saber, como la medicina, la ingeniería, la estadística, la química, la física, la aeronáutica, los viajes espaciales, el sector financiero y el sector educativo, entre otros. La

existencia de *software* como R, Matlab, Fortran, SPSS y SAS, de lenguajes de programación de quinta generación y demás invita a que cada día más investigadores se involucren de alguna manera con la realización de procesos de simulación.

Para desarrollar procesos de simulación es conveniente iniciar por revisar los conceptos involucrados, los procedimientos básicos, los algoritmos para generar valores de variables aleatorias y la escritura de los programas de computador que se requieren para tal fin; posteriormente, será necesario formular modelos con base en variables que permitan comprender de manera aproximada un determinado sistema o fenómeno de interés que se quiera estudiar.

El presente libro es resultado del conocimiento estadístico-probabilístico y de simulación logrado por los autores en su actividad académica como profesores universitarios, y de un proceso de indagación y acrecentamiento de tal conocimiento por medio del desarrollo de un proyecto de investigación. En este trabajo se elabora un marco conceptual sobre los elementos que involucra la simulación con modelos aleatorios, se construyen algoritmos para simular valores de variables aleatorias discretas y continuas usuales o provenientes de la Distribución Lambda Generalizada (DLG), se aporta un algoritmo general para simular variables con DLG, y con los resultados de los procesos de simulación efectuados se generan las tablas que aparecen en cada capítulo; estas tablas y las figuras (gráficas) fueron elaboradas por los autores usando el software libre R. Los mencionados procesos se desarrollaron tomando como base el concepto de número aleatorio.

El texto está dividido en cinco capítulos. En el primero se presentan los elementos conceptuales sobre simulación, que incluye una conceptualización sobre simulación y nociones sobre lo aleatorio, la generación de números aleatorios por distintos métodos, las pruebas estadísticas, la planeación de experimentos de simulación en computadoras y las formas de utilizar la simulación. En el segundo, se describe el conocimiento probabilístico usual en simulación, conformado por los conceptos de variable aleatoria, función de distribución y sus principales momentos; se explicitan de forma amplia los fundamentos teóricos de cinco de los métodos de simulación más utilizados, y se trata el método de Monte Carlo; lo anterior, se hace tomando como base la definición de espacios de probabilidad. En el tercero se indican en forma integrada ciertos modelos de probabilidad discretos y continuos para la simulación de valores de

variables aleatorias, los algoritmos correspondientes, ejemplos de la generación de valores aleatorios en cada caso de forma manual o por computador y se construyen tablas para tal fin. En el cuarto se proporcionan 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); se cierra este capítulo con ejemplos de aplicación. En el quinto se define y caracteriza a la DLG, con sus momentos, y se proporcionan algunos ejemplos de densidades provenientes de esta, incluyendo las gráficas; luego, se desarrollan procesos de simulación para generar valores de variables aleatorias con distribución simétrica y asimétrica provenientes de la DLG mediante un algoritmo general construido con base en el método de la transformada inversa; para finalizar este capítulo se presentan ejemplos de aplicación de la simulación en contextos reales. Al final de cada capítulo se proponen ejercicios que invitan al lector al afianzamiento de sus conocimientos, a la aplicación de los conceptos y al desarrollo de procesos de simulación de valores de variables aleatorias usuales y no usuales.

Para la comprensión del texto, el lector debe tener conocimientos básicos sobre probabilidad y estadística, aritmética, conjuntos, cálculo diferencial e integral y programación de computadores. Es posible que el presente documento tenga errores, se solicita a los lectores hacerlos conocer a los autores dentro de un ambiente académico crítico que permita corregirlos de manera pertinente y oportuna.

1

ELEMENTOS CONCEPTUALES SOBRE SIMULACIÓN

Para desarrollar procesos de simulación se requiere, en primera instancia, acceder a los elementos conceptuales asociados con ellos. En este capítulo se presenta la primera parte de lo que puede llamarse un marco conceptual para la simulación; incluye una conceptualización acerca de simulación, además de las nociones sobre lo aleatorio y la generación de números aleatorios por distintos métodos, con sus correspondientes pruebas estadísticas; sobre la planeación de experimentos de simulación en computadoras, considerando sus etapas (fases), y sobre las formas de utilizar la simulación.

1.1 Concepto de simulación

El concepto de simulación ha evolucionado de forma paulatina desde hace mucho tiempo. Se intuye que la simulación tuvo sus inicios en el momento mismo en que la humanidad empezó a emular los sucesos acaecidos en la naturaleza de manera aleatoria o a imitar el comportamiento de animales y vegetales que desarrollaron procesos muy elaborados para subsistir en determinados ambientes; desde entonces, la humanidad ha elaborado máquinas y diseñado procedimientos que imitan el desplazamiento, el movimiento o las formas de funcionamiento de ellos. De manera formal, la simulación se ha venido trabajando desde la década de los años cuarenta del siglo pasado, cuando Turing inventó una máquina ideal, que funcionaba perfectamente en el papel e imitaba a la computadora actual (Turing, 1950). Usando el método de Monte Carlo se imitaron las explosiones nucleares, trabajando sobre modelos matemáticos (Gentle, 1998).

En términos generales, la simulación puede entenderse como el arte de construir modelos para estudiar el comportamiento de un sistema real. Un modelo es una representación de un sistema de interés a través de variables aleatorias, variables controlables y no controlables. Muchas simulaciones se realizan con base en modelos matemáticos, que

se pueden clasificar en *probabilísticos* y *determinísticos*. La naturaleza del problema por resolver o del sistema por estudiar indica usualmente cuál es el modelo más pertinente para su solución o para su análisis.

En general, en los modelos determinísticos se generan relaciones funcionales (por medio de variables), para las cuales los parámetros del modelo se conocen con certidumbre, pueden regirse por leyes conocidas y producir soluciones únicas; en cambio, en los modelos probabilísticos los parámetros se comportan de manera aleatoria. En concordancia con Blanco (2004), cuando el modelo pertinente es el probabilístico resulta conveniente considerar una o varias variables aleatorias que permitan describir o inferir sobre el problema o sistema en estudio; sin embargo, muchos problemas pueden requerir soluciones que impliquen usar técnicas numéricas para conducir experimentos usando una computadora digital; en este caso la simulación requiere de cierto tipo de modelos matemáticos y lógicos que describan el comportamiento de un sistema o subsistema en determinados periodos del tiempo (generalmente en periodos extensos), usando para ello números aleatorios o técnicas numéricas específicas.

Actualmente, la simulación por computador ha alcanzado grandes desarrollos, y su uso está creciendo muy rápidamente, puesto que se ha constituido en una herramienta poderosa para resolver gran variedad de problemas que por métodos analíticos no se habían podido solucionar en diversos campos del conocimiento humano, tales como estadística, economía, física, biología, ecología, informática e ingeniería, entre otros.

Diversos autores han conceptualizado acerca de la simulación, entre ellos: West Churchman (1973), Naylor (1977), Papoulis (1991), Azarang (1996), Ross (1998), Rios Insúa (2000), Blanco (2004) y Albert (2009), y dejan ver en sus escritos que para realizar procesos de simulación se debe partir de modelos matemáticos que intenten, razonablemente, emular el comportamiento de un sistema de interés; para implementar dichos modelos será necesario utilizar herramientas teóricas provenientes de la estadística matemática, de la teoría de probabilidad, de las ecuaciones diferenciales y de la programación de computadores, entre otras.

Para los autores mencionados, simular es experimentar con modelos matemáticos que describen un sistema de interés utilizando una computadora. Para West Churchman (1973): “ X simula a Y si: X e Y son sistemas formales, Y se considera como un sistema real, X se toma como una aproximación de Y , el modelo X con sus reglas de validez no está exento de error”. En Naylor (1977) se afirma que la simulación es una técnica numérica que permite realizar experimentos en una computadora digital usando modelos matemáticos y lógicos que describen de manera aproximada un sistema de interés.

Agregan los citados autores que para la planeación de experimentos de simulación se requiere de procedimientos que pueden incluir al menos las siguientes etapas:

identificación o formulación del problema, planteamiento de hipótesis, recolección y procesamiento de datos tomados de la realidad –correspondientes a variables controlables o no controlables (aleatorias)–, formulación de modelos matemáticos que relacionen las variables, estimación de parámetros a partir de datos reales (muestras), evaluación de los modelos matemáticos propuestos con base en los parámetros estimados, elaboración de un programa para usarse en una computadora, validación, ejecución del experimento de simulación y análisis de los datos simulados.

Así mismo, siguiendo a dichos autores se puede afirmar que la base fundamental para realizar procesos de simulación son los denominados números aleatorios; estos son utilizados para analizar el comportamiento estocástico de un sistema bajo estudio. Los números aleatorios son aquellos que se comportan caprichosamente como valores de variables aleatorias, no controlables. La idea al generar números aleatorios es que no se pueda determinar el predecesor (mediante alguna fórmula) del siguiente número aleatorio generado. Inicialmente se generaron números aleatorios reales o verdaderos utilizando mecanismos como ruletas y dados; luego, usando métodos como el de cuadrado medio y otros basados en congruencias lineales, se obtuvieron secuencias de números llamados cuasialeatorios, que también se pueden generar por medio del computador. Actualmente, se utilizan métodos computacionales para generar números aleatorios. Los números generados por computadora se denominan pseudoaleatorios, y requieren superar algunas pruebas estadísticas para ser considerados verdaderos números aleatorios.

En Naylor (1977) se sugieren ideas para simular variables tipo Bernoulli y Binomial, y en Blanco (2004) se presentan algunos algoritmos para generar variables aleatorias provenientes de ciertas distribuciones discretas usuales y se indican algunos procesos de simulación para el mencionado tipo de variables. Para variables aleatorias provenientes de distribuciones continuas, caracterizadas por ser simétricas, se han propuesto métodos generales como la transformada inversa, rechazo, convolución, Box-Muller y métodos mixtos, entre otros.

En particular, el tema de la simulación de variables aleatorias normales ha sido tratado por varios autores en distintos contextos: Papoulis (1991) indica un algoritmo para simular una variable aleatoria normal utilizando el método de Box-Muller; Ross (1999) presenta desarrollos teóricos basados en coordenadas polares tendientes a la simulación de una variable normal; Ríos (2000) hace referencia al método de las 12 uniformes para generar una variable con distribución normal; Blanco (2004) diseña un algoritmo para generar valores de una variable aleatoria normal; Burbano (2010), en su artículo publicado en la revista de la Facultad de Ciencias de la Universidad Pedagógica y Tecnológica de Colombia, desarrolla un procedimiento de simulación basado en el método de la transformada inversa, usando la función inversa de la Distribución Lambda Generalizada, citada por Karian y Dudewicz (2000), para simular valores de variables aleatorias con distribución normal, uniforme y logística.

En los textos clásicos de métodos numéricos se proporciona gran variedad de técnicas que pueden ser utilizadas para modelar algunas componentes de un sistema real, en particular, cuando se involucren variables controlables o de tipo determinístico. Por otra parte, los modelos de regresión también pueden contribuir en la formulación de modelos para estudiar relaciones de tipo causal. Así mismo, la estadística matemática aporta elementos teóricos, especialmente para cuando se tengan que realizar procesos tendientes a estimar los parámetros utilizando simulación Monte Carlo.

1.2 Noción de lo aleatorio y de los números aleatorios

El intento por entender y explicar los fenómenos que ocurren de manera inesperada, incierta, caprichosa o azarosa en la naturaleza o en la sociedad ha llevado a la necesidad de buscar herramientas conceptuales para estudiar los fenómenos aleatorios y a utilizar la teoría de la probabilidad como generadora de una medida de la posibilidad de que dichos fenómenos puedan ocurrir; para ello se consideran distintos contextos, que involucran desde las concepciones de tipo intuitivo hasta las que incluyen elementos formales, como espacios de probabilidad y variables aleatorias que permiten, bajo ciertas reglas, calcular probabilidades.

Para modelar situaciones reales en las cuales haya presencia de incertidumbre es conveniente utilizar el razonamiento aleatorio o probabilístico y buscar procedimientos que permitan de manera razonable emular o imitar dichas situaciones. Según Godino, Batanero y Cañizares (1987), la noción de lo aleatorio ha estado ligada a las diferentes concepciones desarrolladas en torno a la probabilidad, que es entendida como la medida de la posibilidad de que un suceso ocurra. Un primer significado sobre lo aleatorio se puede extraer del diccionario de M. Moliner (1983): “Incierto. Se dice de aquello que depende de la suerte o del azar”, siendo el azar “la supuesta causa de los sucesos no debidos a una necesidad natural ni a una intervención humana ni divina”. El anterior contexto permite afirmar que aleatorio es aquello cuyas causas son desconocidas, y que el azar correspondería a la causa de los fenómenos aleatorios. El azar puede suprimir la influencia de la voluntad o del conocimiento de los individuos sobre algo que ocurre de manera fortuita. De acuerdo con Poincaré (1936), los filósofos clásicos ya distinguían los fenómenos aleatorios de los no aleatorios. Los fenómenos aleatorios no podían preverse o determinarse, porque se rebelaban a toda ley, mientras que los no aleatorios parecían obedecer a las leyes conocidas.

La utilización de modelos en contextos concretos constituye un paso necesario para la construcción o para el aprendizaje del conocimiento científico; en la enseñanza del tema de la probabilidad pueden convertirse en un instrumento bastante atractivo. Heitele (1975) incluye la simulación a través de modelos aleatorios entre su lista de ideas estocásticas fundamentales. Según Batanero (2001), la simulación se puede utilizar como un modelo pseudoconcreto en distintas situaciones reales, puesto que

ofrece la posibilidad de trabajar sin los formalismos matemáticos en los casos en que se consideran y se analizan situaciones aleatorias; en las ideas de esta autora, la simulación desempeña un papel intermedio entre la realidad y el modelo matemático; como herramienta matemática puede ser útil para mejorar las *intuiciones* probabilísticas acerca de un fenómeno aleatorio, para elaborar los diferentes pasos (etapas) en el trabajo de modelación de dicho fenómeno y para favorecer la discriminación entre modelo y realidad.

Para realizar procesos de simulación por computador, la base fundamental son los llamados números pseudoaleatorio, que se asemejan a los valores de variables aleatorias, y una vez hayan superado ciertas pruebas se pueden considerar verdaderos números aleatorios, que se comportan de forma similar a los fenómenos no controlables en la naturaleza o las situaciones inciertas en la sociedad.

1.3 Generación de números aleatorios

Se han ideado diversos mecanismos y métodos para obtener secuencias de números que puedan considerarse aleatorios. Inicialmente se generaron números aleatorios usando técnicas manuales como ruedas giratorias o ruletas, lanzamiento de dados o barajas (Ross, 1999). Un generador de números es una fórmula recursiva específica que produce números aleatorios de manera determinística (Blanco, 2004); algunos tipos de generadores se basan en el método de los cuadrados medios o en el método de las congruencias lineales, entre otros, desarrollados de forma manual. Hoy, con frecuencia se usan generadores fundamentados en congruencias lineales e implementados mediante algoritmos computacionales para producir números aleatorios de manera más rápida y eficiente. A continuación se describen algunos de estos métodos.

1.3.1 Método de los cuadrados medios

El *método de los cuadrados medios*, ideado por Neumann en la década de los cuarenta del siglo xx, fue el primer algoritmo de cómputo manual para obtener números pseudoaleatorios. El método se puede resumir de la siguiente forma: se inicia con un número entero positivo de n dígitos, llamado “semilla”, con n par; se toman los k dígitos a la izquierda del dígito $d_{n/2}$, incluyendo a este dígito con $k < n/2$, y los k dígitos a la derecha respetando el orden; los dígitos tomados conforman un nuevo número que se eleva al cuadrado y cuyo resultado se denota con X_1 ; el primer número pseudoaleatorio se construye como un número que pertenece al intervalo $(0,1)$ de la siguiente manera: $U_1 = 0.X_1$; el proceso descrito anteriormente se repite para generar a X_2 , y con este se obtiene $U_2 = 0.X_2$, que corresponde al segundo número pseudoaleatorio; se continúa de similar modo para obtener X_3 , y con este, $U_3 = 0.X_3$, el tercer número pseudoaleatorio, y así sucesivamente hasta llegar a $U_m = 0.X_m$; los anteriores corresponden a números pseudoaleatorios, porque no son resultado del azar.

A manera de ejemplo, a continuación se obtienen 12 números pseudoaleatorios utilizando el algoritmo de Neumann. Se inicia el procedimiento tomando el número 185397 como semilla; aquí $n = 6$, se tiene que $k = 2$, el cual es menor que $3 = n/2$. El dígito $d_{n/2}$ es igual a 5; tomando los dos dígitos a la izquierda de $d_{n/2}$, incluido el 5, y los dos dígitos a la derecha de $d_{n/2}$, resulta el número 8539, que al elevarlo al cuadrado resulta $X_1 = 72914521$, así, $U_1 = 0.72914521$; repitiendo el procedimiento, tomando ahora como base el número $X_1 = 72914521$, se obtiene el número 9145, que al elevarlo al cuadrado da el número $X_2 = 83631025$, del cual resulta $U_2 = 0.83631025$; continuando con el procedimiento se completa la generación de 12 números aleatorios.

$$(8539)^2 = 72914521 \rightarrow 9145$$

$$(9145)^2 = 83631025 \rightarrow 6310$$

$$(6310)^2 = 39816100 \rightarrow 8161$$

$$(8161)^2 = 66601921 \rightarrow 6019$$

$$(6019)^2 = 36228361 \rightarrow 2283$$

$$(2283)^2 = 52120890 \rightarrow 1208$$

$$(1208)^2 = 14592640 \rightarrow 5926$$

$$(5926)^2 = 35117476 \rightarrow 1174$$

$$(1174)^2 = 13782760 \rightarrow 7827$$

$$(7827)^2 = 61261929 \rightarrow 2619$$

$$(2619)^2 = 68591610 \rightarrow 5916$$

$$(5916)^2 = 34999056 \rightarrow 9990$$

En consecuencia, los números aleatorios restantes son: $U_3 = 0.39816100$, $U_4 = 0.66601921$, $U_5 = 0.36228361$, $U_6 = 0.52120890$, $U_7 = 0.14592640$, $U_8 = 0.35117476$, $U_9 = 0.13782760$, $U_{10} = 0.61261929$, $U_{11} = 0.68591610$, $U_{12} = 0.34999056$.

En el método de los cuadrados medios, cuando en la generación de números pseudoaleatorios se obtengan números impares se aumenta un cero a la derecha para tener un número par de dígitos y continuar con el algoritmo. En el caso de que en el proceso se obtengan números pseudoaleatorios repetidos se procede a desecharlos. Cuando en el proceso se obtengan únicamente ceros (como sucede si se continuara con el anterior ejemplo), se procede a tomar otra semilla de forma aleatoria y se continúa con el proceso hasta completar la cantidad de números pseudoaleatorios que se requieran. Luego, a estos números se les deben aplicar pruebas estadísticas

que garanticen su validez para configurarse en un conjunto de números aleatorios que puedan ser utilizados satisfactoriamente en un proceso de simulación.

Los números pseudoaleatorios generados mediante un determinado algoritmo deben presentar buenas propiedades estadísticas, como ajustarse a una distribución uniforme en el intervalo $(0,1)$ y no estar correlacionados entre sí, entre otras. Es conveniente que los algoritmos permitan generar gran cantidad de números pseudoaleatorios antes de que comiencen a repetirse. El algoritmo de Neumann permite generar pequeñas cantidades de números pseudoaleatorios, es decir, muy rápido se converge a números cuyos dígitos son todos ceros, teniéndose que elegir una nueva semilla; este inconveniente se superó al finalizar el siglo xx con la implementación de otros métodos algorítmicos basados en programas de computador.

1.3.2 Método de las congruencias lineales

Muchos algoritmos para generar números aleatorios en $(0,1)$ se basan en el concepto de congruencia lineal. Una congruencia lineal es una relación del tipo $x = (at + b) \bmod (m)$, donde x , t , a , b y m son números enteros; el número x es el residuo que resulta de la división de $at+b$ entre m . El primero en introducir las congruencias lineales en la generación de números pseudoaleatorios fue Lehmer (1951), a mediados del siglo pasado. El generador congruencial se basa en un algoritmo conformado de la siguiente manera: dados los enteros positivos a , b y m , se pueden generar los números pseudoaleatorios U_i en $[0,1)$ con $i = 1, 2, 3, \dots, m$, mediante los pasos que a continuación se indican (Law, 1991):

1. Se escoge aleatoriamente una semilla X_0
2. $X_{i+1} = (aX_i + b) \bmod (m)$
3. $U_i = (X_{i+1})/m$

Con X_0 como un número llamado semilla que cumple la restricción de ser mayor o igual que cero y menor que m ; a , una constante multiplicativa, y b , la constante aditiva. Del segundo paso se deduce que X_{i+1} corresponde al residuo que se obtiene de dividir la expresión $aX_i + b$ entre m . Además, como $X_{i+1} < m$, se deduce que U_i se encuentra en el intervalo $[0,1)$.

Por ejemplo, para la semilla $X_0 = 5$, con $a = 1$, $b = 9$, $m = 14$, se producen los 15 valores para X_i y, por consiguiente, los 15 números pseudoaleatorios U_i ; estos se presentan en la Tabla 1.1, en la cual se visualiza que la congruencia lineal planteada permite obtener máximo 13 números pseudoaleatorios, y luego empieza a repetirlos (los dos últimos coinciden con los dos primeros). A esta repetición, que empieza en U_{14} e iría hasta U_{26} , se le llama el ciclo de la congruencia lineal. Dicho ciclo continúa repitiéndose

infinidad de veces, luego los primeros trece números pseudoaleatorios conforman una clase de equivalencia. Solamente si se seleccionan adecuadamente las constantes enteras y positivas a , b y m , el algoritmo alcanza un ciclo máximo de tamaño igual a m . Si se trabaja con computadoras de 32 bits, una buena selección de n es $m = 2^{31}-1$, $a = 7^5 = 16807$, y $b = 0$ (Saavedra e Ibarra, 2008; Barrera, 2008).

Tabla 1.1. Números aleatorios generados mediante congruencia lineal

X_i	U_i
1	0,0769231
10	0,7692308
6	0,4615385
2	0,1538462
11	0,8461538
7	0,5384615
3	0,2307692
12	0,9230769
8	0,6153846
4	0,3076923
0	0
9	0,6923077
5	0,3846154
1	0,0769231
10	0,7692308

En el anterior contexto, no es suficiente con buscar generadores que tengan ciclo máximo, también es deseable que los números pseudoaleatorios tengan buenas propiedades estadísticas, sean fáciles de generar y de almacenar; además, que la semilla replique la secuencia de los citados números. Un buen generador de números pseudoaleatorios en el intervalo $(0,1)$ es el que se consigue con el procedimiento de Kobayashi, el cual se expresa a través de $X_i = (314159269X_{i-1} + 453806245)$ módulo (2^{31}) (Glasserman, 2004). Hoy, la búsqueda de generadores de números pseudoaleatorios con mejores propiedades y más eficientes sigue abierta a la investigación. Los paquetes estadísticos comerciales incluyen rutinas que permiten generar automáticamente secuencias de números pseudoaleatorios a partir de determinadas semillas.

1.4 Pruebas estadísticas para números pseudoaleatorios

Como se ha anotado, los números aleatorios obtenidos mediante el computador o una calculadora son pseudoaleatorios; para que sean aleatorios y válidos para generar

variables aleatorias deben cumplir ciertas características. Si ellos pertenecen al intervalo $(0,1)$, hipotéticamente provienen de la distribución uniforme en $(0,1)$ y deberían ajustarse a dicha distribución, que presenta una media igual a $1/2$ y una varianza igual a $1/12$; además, ser estadísticamente independientes (no estar correlacionados). Por lo tanto, es recomendable hacerles algunas pruebas estadísticas, entre ellas: la prueba de media, la prueba de varianza y la prueba de independencia. En este sentido, una prueba de hipótesis es un procedimiento que se desarrolla con el propósito de rechazar o no determinada hipótesis (afirmación que se hace acerca de un parámetro) denominada H_0 ; para ello se utiliza un modelo de probabilidad y un nivel de significancia denotado con α . Este α corresponde a la probabilidad de rechazar H_0 cuando sea verdadera; $1-\alpha$ recibe el nombre de nivel de confianza o, simplemente, confiabilidad. Un modelo probabilístico es un modelo matemático que establece una relación entre una o más variables aleatorias y sus correspondientes parámetros (Mayorga, 2003). Los parámetros son valores verdaderos y únicos válidos para toda la población involucrada en un determinado estudio o investigación.

1.4.1 Prueba de la media

Para la *prueba de la media* se sugieren varios caminos:

1. Realizar una prueba de las hipótesis para la media:

$$\begin{aligned} H_0 &: \mu = 1/2 \\ H_1 &: \mu \neq 1/2 \end{aligned}$$

2. Construir un intervalo de confianza para estimar la media poblacional, asumiendo que:

$$\bar{X} = \sum_{i=1}^n \frac{U_i}{n}$$

donde se supone que los U_i son números pseudoaleatorios provenientes de la distribución uniforme en el intervalo $(0,1)$.

Para muestras grandes, la estadística

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}}$$

sigue distribución normal estándar, siempre y cuando los U_i sean independientes e idénticamente distribuidos con los parámetros de la distribución uniforme $E(U_i) = 1/2$ y $\text{Var}(U_i) = 1/12$, entonces,

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

$$Var(\bar{X}) = Var\left(\sum_{i=1}^n \frac{U_i}{n}\right) = \frac{1}{n^2} \left(\sum_{i=1}^n Var(U_i)\right) = \frac{n}{n^2} Var(U_i) = \frac{1}{n} \left(\frac{1}{12}\right) = \frac{1}{12n}$$

Así,

$$Z = \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}}$$

Para el caso 1. se puede utilizar como estadística de prueba la anterior expresión que se puede escribir de la siguiente forma:

$$Z = \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}}$$

Ahora, si la cantidad de números pseudoaleatorios es pequeña ($n < 30$) el estadístico de prueba correspondiente se distribuye como una t-student con $n-1$ grados de libertad dada por

$$T = \frac{\bar{X} - \frac{1}{2}}{\frac{S}{\sqrt{n}}}$$

Para el caso 2. un intervalo del $(1-\alpha)\%$ de confianza para estimar la media poblacional es

$$\mu \in \left(\bar{X} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}, \bar{X} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} \right)$$

Lo anterior significa que si el intervalo contiene a la media $\frac{1}{2} = 0.5$, entonces los números pseudoaleatorios se ajustan a una distribución uniforme en el intervalo $(0,1)$.

Otra forma razonable de analizar si los números aleatorios provienen de la distribución uniforme en $(0,1)$ con una confiabilidad $(1-\alpha)\%$ se tiene que cumplir que

$$P\left(Z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}} \leq Z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$$

Porque hipotéticamente $\mu = \frac{1}{2}$, la anterior expresión se puede escribir como

$$P\left(\frac{1}{2} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{12n}} \leq \bar{X} \leq \frac{1}{2} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} \right) = 1-\alpha$$

Puesto que en la distribución normal estándar $Z_{\frac{\alpha}{2}} = -Z_{1-\frac{\alpha}{2}}$, por tratarse de una distribución simétrica, resulta que

$$P\left(\frac{1}{2} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} \leq \bar{X} \leq \frac{1}{2} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}\right) = 1 - \alpha$$

La anterior expresión indica que si \bar{X} cae dentro del intervalo (L_i, L_s) , los números aleatorios cumplen con la prueba de la media, de lo contrario no sirven y hay que generar unos nuevos. En este caso,

$$L_i = \frac{1}{2} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}, \quad L_s = \frac{1}{2} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}$$

Finalmente, a la muestra de números pseudoaleatorios se les puede aplicar la prueba K-S de Kolmogorov-Smirnov para analizar si se ajustan a una distribución uniforme en el intervalo $(0,1)$.

A manera de ejemplo, a continuación se presentan 40 números pseudoaleatorios generados por el método de Neumann; se desea averiguar si se ajustan a una distribución uniforme en el intervalo $(0,1)$ usando un nivel de confianza del 95%.

0,6101 0,9540 0,2369 0,3385 0,0528 0,7383 0,1198 0,0589
 0,5640 0,1060 0,6850 0,0960 0,9880 0,4500 0,7570 0,2130
 0,7290 0,3670 0,4170 0,9170 0,7670 0,4600 0,5950 0,1080
 0,6560 0,5950 0,6990 0,4930 0,1006 0,7326 0,2254 0,2214
 0,1880 0,0996 0,6759 0,6927 0,8668 0,3512 0,4964 0,7020

Nivel de confianza: $1 - \alpha = 0.95$; $\alpha = 0.05$; $\bar{X} = 0,478$

Al aplicar la prueba de Kolmogorov-Smirnov para los 40 números anteriores, se tiene que la prueba resulta no significativa, es decir, siguen una distribución uniforme de media 0.5 en el intervalo $(0,1)$.

Para verificar las hipótesis:

$$H_0 : \mu = \frac{1}{2}$$

$$H_1 : \mu \neq \frac{1}{2}$$

Se usa una prueba bilateral, como se indica en la Figura 1.1:

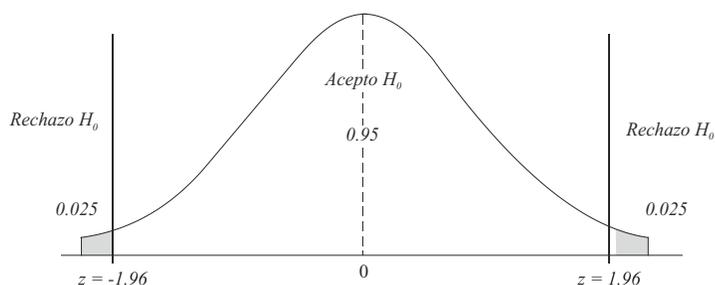


Figura 1.1. Gráfica de la prueba bilateral

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

La estadística de prueba presenta el siguiente resultado:

$$Z = \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}} = \frac{0.478 - \frac{1}{2}}{\frac{1}{\sqrt{12 \cdot 40}}} = \frac{-0.022}{0.0456} = -0.4385$$

Como el valor de la estadística de prueba cae en la región de aceptación de H_0 , se concluye que los números pseudoaleatorios considerados sí tienen una media igual a $\frac{1}{2}$.

Aplicando la prueba t-student, cuya dirección de la prueba se presenta en la Figura 1.2, se tiene:

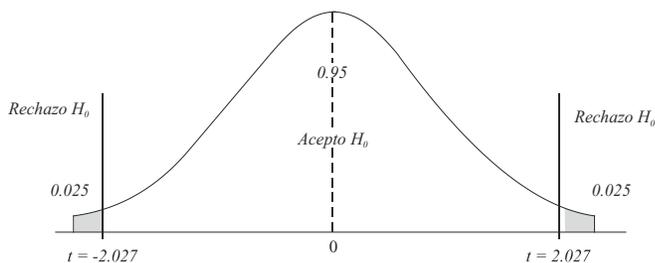


Figura 1.2. Prueba t-student bilateral

(Esta gráfica de la distribución t-student fue elaborada por los autores con el software libre R)

La estadística de prueba, con $S = 0.28127$, permite obtener el siguiente resultado:

$$T = \frac{\bar{X} - \frac{1}{2}}{\frac{S}{\sqrt{n}}} = \frac{0.478 - \frac{1}{2}}{\frac{0.28127}{\sqrt{40}}} = \frac{-0.022}{0.04447} = -0.4947$$

Debido a que el valor de la estadística de prueba cae en la región de no rechazo de H_0 , se concluye que los números pseudoaleatorios considerados sí tienen una media igual a $\frac{1}{2}$.

El intervalo de confianza del 95% es el siguiente:

$$\mu \in \left(\bar{X} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}, \bar{X} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} \right)$$

$$\mu \in \left(0.478 - \frac{1.96}{\sqrt{12(40)}}, 0.478 + \frac{1.96}{\sqrt{12(40)}} \right)$$

$$\mu \in (0.3886, 0.5674)$$

Como el intervalo anterior contiene el valor $\frac{1}{2}$, se deduce que los números pseudoaleatorios tienen su media estadísticamente igual a $\frac{1}{2}$.

Al obtener el intervalo (L_i, L_s) resulta:

$$L_i = \frac{1}{2} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} = \frac{1}{2} - \frac{1.96}{\sqrt{12 * 40}} = 0.5 - 0.0894 = 0.4106$$

$$L_s = \frac{1}{2} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} = \frac{1}{2} + \frac{1.96}{\sqrt{12 * 40}} = 0.5 + 0.0894 = 0.5894$$

Debido a que el intervalo $(0.4106, 0.5894)$ contiene la media poblacional, se concluye que los 40 números pseudoaleatorios superan la prueba de la media.

En la práctica, es suficiente con aplicar una de las pruebas citadas anteriormente.

1.4.2 Prueba de la varianza

De acuerdo con Azarang (1996), para la prueba de la varianza se calcula la varianza de los números pseudoaleatorios generados, mediante la expresión

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

donde los números r_i pertenecen a una variable U_i .

La variable aleatoria siguiente:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

sigue una distribución chi cuadrado con $n - 1$ grados de libertad. Como se dispone del valor $\sigma = \sqrt{\frac{1}{12}}$, se puede obtener un intervalo de confianza que contenga a S^2 ; a dicho intervalo se lo denota con (L_i, L_s) ; donde:

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

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

Para los siguientes 10 números pseudoaleatorios, aplicar la prueba de la varianza:

0.685, 0.676, 0.028, 0.858, 0.788, 0.077, 0.068, 0.318, 0.906, 0.533

Para: $\alpha = 0.05$

$$\bar{X} = 0,4937$$

$$S^2 = 0,1181$$

Los límites inferior y superior son:

$$L_i = \frac{\chi_{0.025}^2}{12(9)} = \frac{2.7}{108} = 0.025$$

$$L_s = \frac{\chi_{0.95}^2}{12(9)} = \frac{19}{108} = 0.1759$$

El intervalo (0.025,0.1759) contiene la varianza poblacional y el valor $S^2 = 0,1181$; luego los 10 números pseudoaleatorios sí cumplen con la prueba de varianza.

1.4.3 Prueba de independencia

La *prueba de independencia* permite verificar que los números pseudoaleatorios no estén correlacionados. Existen varias pruebas para verificar independencia, entre ellas están la prueba de Póker, de corridas (rachas), de distancias, de series y de huecos, entre otras. A continuación se describe la prueba de corridas (rachas); el proceso para aplicar la prueba de rachas arriba y abajo involucra los siguientes pasos:

1. Generar números pseudoaleatorios
2. Clasificar cada número pseudoaleatorio teniendo en cuenta el siguiente criterio:

Si $r_i < r_{i-1}$, se pone un signo menos (-)

Si $r_i > r_{i-1}$, se pone un signo más (+)

Para $i = 2, 3, \dots, n$. Los r_i son números pseudoaleatorios

Calcular el número de corridas (rachas) observado h ; el número de corridas es una variable aleatoria, con:

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

3. Calcular el valor esperado y la varianza de la variable aleatoria, por medio de las expresiones anteriores.
4. Aproximar por una variable con distribución normal estándar.

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

Si se cumple la condición $Z < Z_{\frac{\alpha}{2}}$ para un α nivel de significancia, entonces los números pseudoaleatorios son independientes.

Ejemplo: Analizar si los números pseudoaleatorios que se presentan en la Tabla 1.2, son independientes con $\alpha = 0.05$.

Tabla 1.2. Números pseudoaleatorios y sus rachas

Pseudo-aleat.	Signo	Pseudo-aleat.	Signo	Pseudo-aleat.	Signo	Pseudo-aleat.	Signo
0,95572		0,45172	+	0,58647	+	0,27969	-
0,13539	-	0,97453	+	0,40432	+	0,69555	+
0,46635	+	0,09348	-	0,86196	+	0,90257	+
0,30061	-	0,58594	+	0,88295	+	0,13281	-
0,00762	-	0,44389	-	0,60754	-	0,50797	+
0,19391	+	0,34137	-	0,83127	+	0,11783	-
0,89932	+	0,21816	-	0,95661	+	0,92472	+
0,42873	-	0,84197	+	0,19798	-	0,51283	-
0,59252	+	0,14418	-	0,11859	-	0,27146	-
0,78193	+	0,27192	+	0,48467	+	0,64636	+
0,35591	-	0,99225	+	0,19913	-	0,64378	-
0,64527	+	0,72784	-	0,97986	+	0,85229	+
0,43069	-	0,88321	+	0,55882	-	0,06515	-
0,05221	-	0,38914	-	0,68869	+	0,40679	+
0,04221	-	0,37914	-	0,69869	+	0,41679	+

$h = 40$

$n = 60$

$$E(h) = \frac{2(60) - 1}{3} = 39.66$$

$$Var(h) = \frac{16(60) - 29}{90} = 10.34$$

$$Z = \frac{40 - 39.66}{\sqrt{10.34}} = 0.1057$$

En la Figura 1.3 se observa que el valor de la estadística de prueba $Z = 0.1057$ cae en la región de no rechazo de H_0 ; se concluye que los números pseudoaleatorios son independientes.

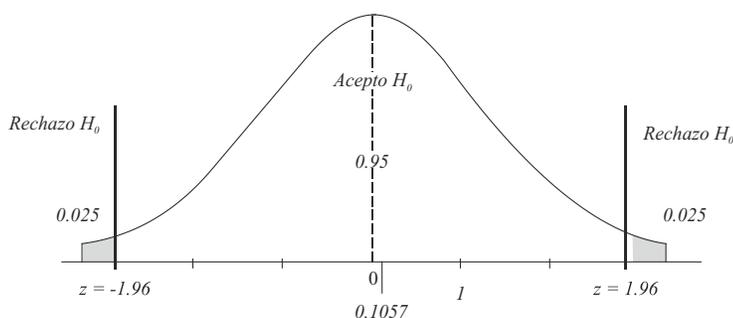


Figura 1.3. Regiones de aceptación y rechazo de H_0
(Esta gráfica de la distribución normal estándar fue elaborada por los autores con el software libre R)

1.5 Planeación de experimentos de simulación en PC

La realización de procesos de simulación involucra la planeación de experimentos de simulación en computadoras (PC), la cual requiere de un procedimiento que puede constar de las siguientes etapas (Naylor, 1993): formulación del problema, planteamiento de hipótesis, recolección y procesamiento de datos tomados de la realidad, formulación del modelo matemático, estimación de parámetros a partir de datos reales, evaluación del modelo y de los parámetros estimados, formulación de un programa para la computadora, validación, diseño de experimento de simulación y análisis de datos simulados. A continuación se presenta una breve descripción de las anteriores etapas.

1.5.1 Formulación del problema

Para modelar un sistema se deben tener en cuenta tres elementos:

Objetivos

Estructura u organización

Entorno o medioambiente

Con esos elementos se debe hacer un diagrama de bloque que puede involucrar tres tipos de variables:

Variables de entrada o exógenas o variables independientes

Variables de estado del sistema

Variables de salida o respuestas o variables endógenas o variables dependientes

La interacción de estas variables se puede visualizar en la Figura 1.4.



Figura 1.4. Relación entre variables de entrada, estado y salida en un sistema

Las variables de salida tienen que ver con el logro (alcance) de los objetivos; cuando estos no se alcanzan, se entiende que hay problemas. La magnitud de un problema puede entenderse como la diferencia entre lo que debería estar ocurriendo (objetivo) y lo que está ocurriendo en la realidad o sistema.

Las variables de entrada están ligadas con la influencia que ejerce el entorno o el medioambiente hacia el sistema. Hay variables o factores que pueden favorecer o entorpecer el logro del objetivo; estos factores o variables pueden ser:

Controlados

No controlados o aleatorios

Las variables no controladas son a las que posiblemente se les tenga que hacer un proceso de simulación a través de modelos aleatorios. Para estas variables se toman datos históricos, y con base en ellos se estiman parámetros que sirven de base para realizar simulaciones. En cambio, es posible que para algunas variables controladas se tengan que implementar modelos matemáticos en los cuales se requiera hacer procesos de cálculo, mediante aproximaciones sucesivas (usando métodos numéricos); es decir, se tenga que realizar simulaciones con modelos no aleatorios.

1.5.2 Recolección y procesamiento de datos tomados de la realidad

Una vez identificado el problema, se debe determinar sobre cuáles variables aleatorias se debe realizar el proceso de simulación, y sobre estas se hace la recolección de los

datos históricos correspondientes. Las variables pueden estar en una determinada escala de medición. Una variable está en *escala nominal* si permite organizar la información (datos) en grupos o clases; en *escala ordinal*, si los valores de la variable permiten establecer un orden; en *escala de intervalo*, cuando sus valores pertenecen a un intervalo de números reales con cero relativo; en *escala de razón*, si la variable admite como posible valor cualquier número real perteneciente a un intervalo dado, con cero absoluto (aquí el cero indica ausencia de la característica de estudio).

Con los datos de la variable es posible realizar pruebas de bondad de ajuste, con el propósito de establecer si la variable aleatoria sigue un modelo teórico de probabilidad, el cual puede ser conocido o desconocido (debe construirse). Entre los *modelos conocidos o usuales* pueden mencionarse los modelos discretos de Bernoulli, binomial, Poisson o uniforme discreto, entre muchos otros; entre los *modelos continuos* están el modelo normal, gamma, Weibull, exponencial, uniforme continuo, la Distribución Lambda generalizada (DLG), etc., modelos para analizar causalidad, como los modelos de regresión lineal (simple, múltiple), modelos de regresión logística, modelos lineales y no lineales, paramétricos y no paramétricos, entre otros.

Si el modelo no resulta apropiado, se implementa uno nuevo o se corrige el que ya se tenía.

1.5.3 Modelo matemático

El modelo matemático está conformado por las relaciones entre las variables de entrada, proceso y salida por medio de funciones o aplicaciones de una o varias variables; en estas se deben incluir las restricciones a que haya lugar, algunas de ellas escritas en forma de igualdades, y otras, en forma de inequaciones. En la construcción del modelo matemático es aconsejable tener presente los siguientes pasos:

1. Especificación de los componentes
2. Especificación de las variables
3. Especificación de las relaciones funcionales

Así mismo, resulta pertinente hacer las siguientes consideraciones:

1. Cuántas variables se deben incluir en un modelo
2. El modelo no debe ser ni tan simple, ni tan complejo
3. Tener en cuenta el tiempo de cómputo; este factor es importante, sobre todo, cuando se tiene la facilidad de los computadores
4. Considerar el tiempo consumido en la programación
5. Determinar la validez del realismo incorporado en el modelo
6. Los tipos de experimentos de simulación por realizar

1.5.4 Estimación de parámetros a partir de datos reales

Cuando se obtienen los datos correspondientes a las variables especificadas, estos se constituyen en datos muestrales, y con base en ellos se obtienen estimadores de los parámetros. También hay necesidad de realizar procesos de inferencia estadística a fin de estimar los parámetros a que haya lugar.

Entre otros, algunos métodos utilizados para estimar los parámetros son:

1. El de mínimos cuadrados
2. El de máxima verosimilitud
3. El de los momentos

A continuación se presenta un ejemplo de estimación por el método de máxima verosimilitud para el parámetro del modelo aleatorio de Bernoulli.

Si X es una variable aleatoria cuya función de probabilidad está dada por:

$$f(x, p) = p^x (1 - p)^{1-x}$$

donde p es el parámetro, y la variable aleatoria X toma los valores 0, 1, entonces dicha variable sigue el modelo de Bernoulli. Para una muestra aleatoria X_1, X_2, \dots, X_n tomada de una población con $f(x, p) = p^x (1 - p)^{1-x}$, la función de verosimilitud es:

$$L(p) = \prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i}$$

donde los x_i son observaciones correspondientes a las variables aleatorias X_1, X_2, \dots, X_n .

Se trata de encontrar un valor del parámetro de tal manera que se maximice la función de verosimilitud

$$L(p) = p^{x_1} (1 - p)^{1-x_1} \cdot p^{x_2} (1 - p)^{1-x_2} \dots p^{x_n} (1 - p)^{1-x_n}.$$

$$L(p) = p^{x_1 + x_2 + \dots + x_n} (1 - p)^{1 + 1 + \dots + 1 - (x_1 + x_2 + \dots + x_n)}$$

$$L(p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

$$\ln(L(p)) = \ln\left(p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}\right)$$

$$\ln(L(p)) = \sum_{i=1}^n x_i \ln(p) + \left(n - \sum_{i=1}^n x_i\right) \ln(1 - p)$$

Ahora, se hace uso del cálculo para derivar parcialmente con respecto al parámetro p

$$\begin{aligned}\frac{\partial}{\partial p}(Ln(L(p))) &= \frac{\partial}{\partial p} \left(\sum_{i=1}^n x_i Ln(p) \right) + \frac{\partial}{\partial p} \left(\left(n - \sum_{i=1}^n x_i \right) Ln(1-p) \right) \\ \frac{\partial}{\partial p}(Ln(L(p))) &= \sum_{i=1}^n x_i \left(\frac{1}{p} \right) + \left(n - \sum_{i=1}^n x_i \right) \left(\frac{-1}{1-p} \right)\end{aligned}$$

Igualando a cero se tiene:

$$\begin{aligned}\frac{\partial}{\partial p}(Ln(L(p))) &= \sum_{i=1}^n x_i \left(\frac{1}{p} \right) + \left(n - \sum_{i=1}^n x_i \right) \left(\frac{-1}{1-p} \right) = 0 \\ \sum_{i=1}^n x_i \left(\frac{1}{p} \right) &= \left(n - \sum_{i=1}^n x_i \right) \left(\frac{1}{1-p} \right) \\ (1-p) \sum_{i=1}^n x_i &= p \left(n - \sum_{i=1}^n x_i \right) \\ \sum_{i=1}^n x_i - p \sum_{i=1}^n x_i &= np - p \sum_{i=1}^n x_i \\ np &= \sum_{i=1}^n x_i \\ p &= \frac{\sum_{i=1}^n x_i}{n} \quad \hat{p} = \frac{\sum_{i=1}^n x_i}{n}\end{aligned}$$

La última expresión corresponde al estimador de máxima verosimilitud del parámetro asociado a la distribución de Bernoulli.

1.5.5 Evaluación del modelo y de los parámetros estimados

Si ya se tienen formulados los modelos matemáticos, se deben valorar, a fin de establecer si son pertinentes y representativos de la realidad que se está modelando. Si se presentan dudas, es el momento de formular nuevamente los modelos, o de seleccionar el mejor modelo mediante un proceso de inclusión o exclusión de algunas variables; una manera de conseguirlo consiste en aplicar métodos estadísticos que permitan determinar la significancia de los parámetros involucrados en el modelo. En este punto conviene contar con un modelo matemático parsimonioso que con el menor número de variables pueda explicar de manera significativa el comportamiento del sistema en consideración.

1.5.6 Formulación de un programa para la computadora (PC)

Si se tiene la seguridad de que el modelo es el adecuado, se debe iniciar el proceso de simulación en la computadora, y para ello es recomendable:

1. Hacer un diagrama de flujo (de datos)
2. Escoger un lenguaje de simulación apropiado, entre ellos: los lenguajes de alto nivel (C++, Fortran, Pascal, Lisp, etc.), los lenguajes especializados como: Simulate, @Risk, GPSS, SPSS, SAS, etc.
3. Realizar búsquedas de errores; los cuales pueden ser de dos tipos: de tipo lógico, en la estructura del programa, o de tipo sintáctico, al digitar u omitir líneas de programación.
4. Determinar datos de entrada e inicialización de variables: para ello se toman datos históricos de lo que ha venido ocurriendo, y con base en ellos se ubican las distribuciones y los parámetros para hacer las simulaciones.
5. Efectuar la generación de datos; una vez se tienen listos los programas, al ejecutarlos ellos dan como salida datos aleatorios con los cuales se trata de hacer análisis estadístico. Aquí se presenta una dificultad, por cuanto los números con los que se generan las simulaciones son pseudoaleatorios, por eso se deben hacer algunas pruebas de aleatoriedad, como la prueba de la media, la prueba de varianza y la prueba de independencia, entre otras.
6. Obtener los reportes de salida; los procesos de simulación dan unos resultados que son confiables siempre y cuando el proceso de simulación se haya efectuado bien.

1.5.7 Validación

En general, el problema de validar modelos de simulación es un tanto complejo, ya que lleva consigo un sinnúmero de situaciones de tipo práctico, teórico, estadístico e inclusive filosófico; sin embargo, se recomienda tener en cuenta:

1. Qué tanto coinciden los datos simulados con los datos históricos
2. Qué tanto se parecen las predicciones del comportamiento del sistema real realizadas con los datos simulados, y las realizadas con los datos muestrales del sistema real, en un momento dado.

1.5.8 Diseño de los experimentos de simulación

Una vez que se tiene un modelo que funciona, al que se le ha efectuado la programación correspondiente (pertinente) y que permite obtener resultados, la parte más difícil de la simulación es concluir si los resultados son adecuados y si las variables independientes explican en alguna medida cierta variable dependiente o si permiten predecir en cierto porcentaje una variable dependiente, o si las variables presentan algún tipo de correlación, asociación o multicolinealidad.

1.5.9 *Análisis de los datos simulados*

Cuando se cuenta con datos reales se pueden aplicar técnicas estadísticas corrientes, pero en el momento que se tienen datos simulados a través de números aleatorios en una computadora el asunto es ya más complejo; sin embargo, una manera de resolver parcialmente el problema es aplicar técnicas estadísticas a los datos simulados, para ello se sugiere seguir tres etapas:

1. Obtención de los datos simulados
2. Cálculo de estadísticas descriptivas a los datos simulados e interpretación de resultados
3. Análisis univariados y multivariados para los datos simulados

1.6 Formas de utilizar el concepto de simulación

En esta sección se describen dos formas de utilizar el concepto de simulación: una que consideramos de nivel intuitivo y otra de nivel avanzado; además, se menciona la necesidad del conocimiento estadístico-probabilístico para el desarrollo de procesos de simulación.

1.6.1 *Nivel intuitivo (elemental)*

En concordancia con Batanero (2001), la simulación estadística consiste en poner en correspondencia dos experimentos aleatorios diferentes, con la condición de que a cada evento elemental del primer experimento le corresponda un suceso elemental del segundo, y solo uno, de tal forma que los sucesos puestos en correspondencia en ambos experimentos tengan igual probabilidad de suceder; este concepto es semejante al de isomorfismo, definido en otras ramas de las matemáticas o en otras ciencias.

En el nivel elemental o intuitivo, un importante uso de la simulación hace referencia a sustituir un experimento aleatorio difícil de observar en la realidad, por otro equivalente. Para ello es conveniente desarrollar un número suficiente de actividades relacionadas con ocurrencias del fenómeno (aumento del tamaño de la muestra). El atractivo que tiene la simulación en este nivel es que se puede efectuar varias veces el segundo experimento y usar sus resultados para obtener información del primero. En Polya (1982) se proporcionan algunos ejemplos de cómo simular procesos aleatorios por medio de una secuencia de extracciones de bolas ubicadas en urnas o recipientes contruidos para tal fin.

1.6.2 Nivel formal (axiomatización, nivel avanzado)

Para desarrollar procesos formales de simulación es conveniente tener presente la conceptualización hecha al respecto por Churchman (1973); además, es necesario hacer una reflexión profunda sobre el fenómeno que se ha de simular, hasta comprenderlo y poder generar un conjunto de pasos bien estructurados que permitan desarrollar un proceso de simulación en el sentido descrito por Naylor (1993), atendiendo la conceptualización formal de Churchman.

1.6.3 Conocimiento estadístico-probabilístico y simulación

El conocimiento estadístico-probabilístico es fundamental para desarrollar procesos de simulación por medio de modelos aleatorios. Algunos autores, como Jones, Langrall y Mooney (2007), consideran la probabilidad como parte de las matemáticas, y hacen énfasis en que es necesario clarificar cuál es el conocimiento del contenido matemático requerido para realizar investigación y cuál es el que se necesita para la enseñanza en los diferentes niveles de cualquier sistema educativo. Para Aliaga y Gunderson (2005), la estadística se ha consolidado como la ciencia de los datos; desde este contexto fundamenta los procesos de simulación.

Para Hill, Ball y Schilling (2008), el conocimiento del contenido matemático incluye el Conocimiento Común del Contenido (CCC), el Conocimiento Especializado del Contenido (CEC) y el Conocimiento en el Horizonte Matemático (CHM). El CCC es el conocimiento activado y utilizado por cualquier persona del común en la resolución de problemas matemáticos. El CEC involucra aspectos del conocimiento matemático considerados no propios de una persona del común para resolver problemas matemáticos (ej. identificar las ideas sobre modelos matemáticos involucrados en problemas de simulación). El CHM es un conocimiento que proporciona perspectivas a los investigadores para su trabajo en relación con otras disciplinas (ej. la simulación). En el contexto anterior, la estadística y la probabilidad tienen un papel fundamental como herramientas que soportan los procesos de simulación realizados a través de variables aleatorias.

Ejercicios capítulo 1

- 1.1 Exponga su propio concepto de simulación. Explique los pasos que implica desarrollar un proceso formal de simulación.
- 1.2 Establezca la diferencia entre lo aleatorio y el azar. Proporcione un concepto de número aleatorio.
- 1.3 Por el método de los cuadrados medios, usando la semilla 924765 genere los números aleatorios que sean posibles. Verificar el cumplimiento de las pruebas estadísticas.
- 1.4 Usando la congruencia lineal $X_{i+1}=(aX_i +b) \bmod (m)$ con $a = 2$, $X_0= 4$, $b=8$, $n=12$ genere los números aleatorios correspondientes y estudie el ciclo de dicha congruencia.
- 1.5 Aplique las pruebas estadísticas pertinentes para verificar si el siguiente conjunto de números pseudoaleatorios, provenientes de una distribución uniforme en el intervalo $[0,1)$, contiene números aleatorios:

0,13964; 0,43130; 0,61218; 0,29075; 0,15573; 0,69950; 0,34630
0,44564; 0,05242; 0,10320; 0,14116; 0,04294; 0,62171; 0,15359
0,71521; 0,92830; 0,57813; 0,26186; 0,72483; 0,03709; 0,10057
0,73264; 0,22541; 0,22142; 0,61009; 0,95397; 0,23690; 0,33852
0,05281; 0,73835; 0,11981; 0,05889; 0,18802; 0,09956; 0,67587
0,69265; 0,86679; 0,35123; 0,49642; 0,70198; 0,20601; 0,77339
0,06692; 0,52707; 0,66184; 0,93580; 0,31569; 0,52272; 0,76585
0,12707.
- 1.6 Proporcione un ejemplo de estimador máximo verosímil para estimar los parámetros del modelo binomial.
- 1.7 Explique en qué consiste la simulación a nivel intuitivo.
- 1.8 Identifique los pasos necesarios para realizar las pruebas estadísticas a secuencias de números pseudoaleatorios.

2

CONOCIMIENTO PROBABILÍSTICO USUAL EN SIMULACIÓN

Diversos procesos de simulación con modelos aleatorios se desarrollan tomando como base los conceptos de variable aleatoria, de función de distribución y de sus principales momentos, entre los que se destacan el valor esperado, la varianza, la asimetría y la curtosis; conceptos que se exponen en este capítulo, como segunda parte del marco conceptual para la simulación; también se presentan los fundamentos teóricos de cinco de los métodos de simulación más utilizados y del método de Monte Carlo. Para formalizar los citados conceptos, se requiere definir elementos como: experimento aleatorio, sigma álgebra (σ – álgebra), medida de probabilidad y espacios de probabilidad, entre otros; estos elementos se han adaptado a este contexto a partir de la literatura existente y se explicitan en las secciones iniciales del presente capítulo.

2.1 Elementos de probabilidad

En esta sección se abordan los conceptos de experimento aleatorio, sigma álgebra y medida de probabilidad; así mismo, se mencionan algunas propiedades de las operaciones con eventos, las cuales posibilitan el cálculo de probabilidades.

2.1.1 Experimento aleatorio

Un experimento aleatorio es aquel cuyos resultados no pueden ser determinados de antemano (Blanco, 2004). Al conjunto de todos los posibles resultados de un experimento aleatorio se le conoce con el nombre de espacio muestral, que suele denotarse con Ω . El espacio muestral Ω se llama *discreto* si Ω es un conjunto finito o numerable, y *continuo* si Ω corresponde a un conjunto infinito no numerable. Un subconjunto T de Ω para el cual sea posible asignar una *medida* numérica de su posibilidad de ocurrencia se denomina *evento* o *suceso*.

2.1.2 Sigma álgebra

Dado un espacio muestral $\Omega \neq \emptyset$. Una familia \mathfrak{T} de subconjuntos de Ω se llama un σ -álgebra sobre Ω , si se cumplen los siguientes axiomas:

1. $\Omega \in \mathfrak{T}$
2. Si $T \in \mathfrak{T}$ entonces $T^c \in \mathfrak{T}$
3. Si $T_1, T_2, \dots \in \mathfrak{T}$ entonces $\bigcup_{i=1}^{\infty} T_i \in \mathfrak{T}$

Si $\Omega \neq \emptyset$ es un espacio muestral y \mathfrak{T} es un σ -álgebra sobre Ω , a la pareja (Ω, \mathfrak{T}) se suele llamar espacio medible. Los subconjunto del espacio medible se denominan eventos.

Un σ -álgebra que se requiere para definir variables aleatorias reales es el σ -álgebra de Borel. Siguiendo a Blanco (2004), el σ -álgebra de Borel es la menor σ -álgebra definida sobre el espacio muestral $\Omega = R$, donde R representa el conjunto de los números reales, la cual contiene a todos los intervalos de la forma $(-\infty, x]$ con $x \in R$; ésta se denota con β .

2.1.3 Medida de probabilidad

En concordancia con Lindgren (1993), dado un espacio medible (Ω, \mathfrak{T}) , una función de conjunto definida desde el σ -álgebra \mathfrak{T} hacia el intervalo $[0,1] \subset R$,

$$P : \mathfrak{T} \rightarrow [0,1]$$

es una medida de probabilidad si satisface los siguientes axiomas:

1. $P(T) \geq 0$ para todo $T \in \mathfrak{T}$
2. $P(\Omega) = 1$
3. Si T_1, T_2, \dots es una sucesión de eventos en \mathfrak{T} tales que $T_i \cap T_j = \emptyset$ para todo $i \neq j$ entonces,

$$P\left(\bigcup_{n=1}^{\infty} T_n\right) = \sum_{n=1}^{\infty} P(T_n)$$

A la función P se le denomina una medida de probabilidad definida sobre el espacio muestral Ω . A la tripleta ordenada $(\Omega, \mathfrak{T}, P)$ se le llama un espacio de probabilidad sobre Ω .

También de Lindgren (1993) se han extractado las siguientes propiedades, que se mencionan en esta sección debido a que en ellas reposa el cálculo de probabilidades.

Sea $(\Omega, \mathfrak{T}, P)$ un espacio de probabilidad sobre Ω , entonces:

1. Para todo $T \in \mathfrak{F}$ se cumple que $P(T^c) = 1 - P(T)$
2. $P(\emptyset) = 0$
3. Para todo $T, S \in \mathfrak{F}$, $P(T \cup S) = P(T) + P(S) - P(T \cap S)$
4. Si $T, S \in \mathfrak{F}$ y $T \subset S$ entonces $P(T) \leq P(S)$
5. Si $P(T) = 0$ entonces $P(T \cap S) = 0$
6. $P(T) = P(T \cap S) + P(T \cap S^c)$
7. Para $T, S \in \mathfrak{F}$ se cumple que,

$$P(T \cap S) \leq P(T \cup S) \leq P(T) + P(S)$$
8. Si $T, S \in \mathfrak{F}$ y $T \subset S$ entonces $P(S - T) = P(S) - P(T)$
9. Si $T_1, \dots, T_n \in \mathfrak{F}$ y $T_i \cap T_j = \emptyset$ para todo $i \neq j$ entonces

$$P\left(\bigcup_{i=1}^n T_i\right) = \sum_{i=1}^n P(T_i)$$

2.2 Variable aleatoria real

Una variable aleatoria real es una aplicación definida desde el espacio muestral hacia el conjunto de los números reales de la siguiente forma:

$$X: \Omega \rightarrow R$$

de modo que para todo evento T en el σ - álgebra de Borel se tiene que

$$X^{-1}(T) \in \mathfrak{F}$$

En general, una variable aleatoria X es una aplicación que tiene la propiedad de ser una función medible. En sí, el espíritu de las variables aleatorias es el de transformar los elementos de un espacio de probabilidad en números reales; esto debido a que en la mayoría de los casos son de interés los valores numéricos que se puedan deducir de un experimento aleatorio enmarcado en un determinado contexto.

En los términos de Valdivieso (2010), dada X , una variable aleatoria definida sobre un espacio muestral Ω , y R_X su rango en R . Si R_X es un conjunto contable (discreto), un conjunto finito o infinito contable, entonces X se denomina variable aleatoria discreta. Si R_X es un subconjunto no contable (infinito no contable) de R , entonces X es una variable aleatoria continua. De manera intuitiva, el rango de una variable aleatoria continua corresponde a un intervalo.

Es pertinente aclarar que los procesos de simulación que incluyen variables aleatorias continuas utilizando números aleatorios provenientes de una distribución uniforme en el intervalo $(0,1)$ implican el uso de procedimientos que trabajan con muestras de las señales continuas; en general, mediante un número grande de iteraciones, estos terminan ejecutándose como procesos computacionales con valores de una variable aleatoria discreta, los cuales simulan los valores de las variables continuas.

Ahora, si $(\Omega, \mathfrak{F}, P)$ es el espacio de probabilidad sobre el cual se define la variable aleatoria X , y (R, β, P_x) es el espacio de probabilidad inducido por la variable aleatoria X , donde β es el σ -álgebra de Borel, y P_x es la medida de probabilidad inducida, para $x \in R$, entonces se tiene que

$$P_x(\{x\}) = P(X^{-1}\{x\})$$

Ahora, si X es una variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, con valores en el espacio medible (R, β, P_x) , y si se define el evento

$$\{X \in S\} = \{\omega \in \Omega: X(\omega) \in S\} \text{ con } S \in \beta$$

entonces

$$P_x(S) = P(\{X \in S\}) \text{ para todo } S \in \beta$$

En el contexto anterior, se puede verificar que P_x cumple con las condiciones que definen una medida de probabilidad.

2.3 Función de probabilidad para una variable aleatoria discreta

Para una variable aleatoria discreta X , definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, con rango R_x (recorrido), una función tal que para cada $x \in R$,

$$f(x) = P_x(\{x\})$$

se denomina la función de probabilidad (*f. p.*) de la variable aleatoria X .

Al conjunto $(\{x, f(x)\}: x \in R)$ se le suele llamar la gráfica de la función de probabilidad de la variable aleatoria X ; esta se puede representar en un plano cartesiano o mediante una tabla de valores, tal como se indica en la Tabla 2.1.

Tabla 2.1. Forma tabular de la función de probabilidad

X	x_1	x_2	\dots	x_n
$f(x) = P(X = x)$	$f(x_1)$	$f(x_2)$	\dots	$f(x_n)$

La función de probabilidad cumple las dos condiciones siguientes:

1. $f(x) \geq 0$
2. $\sum_{x_i \in R_X} f(x_i) = 1$

2.4 Función de densidad de probabilidad

Cada función real f que satisface las dos condiciones siguientes:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$

se denomina función de densidad de probabilidad (*f.d.p.*) para la variable aleatoria continua X . También a f se le llama el modelo de probabilidad que se requiere para desarrollar procesos de simulación de variables aleatorias (Ross, 1998).

Si X es una variable aleatoria tal que para todo evento $S \in \beta$ se cumple que

$$P_X(S) = \int_S f(x) dx$$

para alguna función de densidad f , entonces la variable aleatoria X se llama una variable aleatoria de tipo continua o una variable aleatoria continua. Es de notar que como $S \in \beta$, entonces S es un conjunto Boreliano, y la integral sobre S es una integral de Lebesgue.

2.5 Función de distribución

Se le denomina función de distribución a una función real definida sobre el conjunto de los números reales con valores en el intervalo $[0,1]$, y que satisface los siguientes axiomas:

1. $0 \leq F(x) \leq 1$ para todo $x \in R$
2. Si $x_1 \leq x_2$ entonces $F(x_1) \leq F(x_2)$
3. $F(x^+) = \lim_{a \rightarrow 0^+} F(x+a) = F(x)$
4. $\lim_{x \rightarrow \infty} F(x) = 1$
5. $\lim_{x \rightarrow -\infty} F(x) = 0$

Esta función se caracteriza por ser no decreciente.

Para una variable aleatoria real X definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, la función real F_X tal que para todo $x \in R$ satisface

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x)$$

se le llama función de distribución de la variable aleatoria o función de distribución acumulativa de probabilidad (*f.d.c.*)

Ahora, si X es una variable aleatoria discreta, entonces su función de distribución de probabilidad se define por medio de

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

Si X es una variable aleatoria continua, entonces su función de distribución de probabilidad es

$$F_X(x) = F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Se dice que una variable aleatoria real X definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ es absolutamente continua si y sólo si existe una función real no negativa e integrable f_X , tal que para todo $x \in R$ se tiene que

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

En este caso, la función f_X corresponde a la función de densidad de probabilidad (*f.d.p.*) de la variable aleatoria X , y la función F_X es su función de distribución.

2.6 Esperanza matemática y varianza de una variable aleatoria

Los siguientes conceptos referidos a la esperanza matemática, la varianza y los momentos de una variable aleatoria son adoptados en lo expuesto por Lindgren (1993) y Shao (1999).

Si X es una variable aleatoria discreta definida sobre el espacio de probabilidad $(\Omega, \mathfrak{I}, P)$, entonces el valor esperado de X o esperanza matemática de X está dado por

$$\mu = E(X) = \sum_{x \in R_X} xP(X = x)$$

Si X es una variable aleatoria real continua definida sobre el espacio de probabilidad $(\Omega, \mathfrak{I}, P)$, entonces el valor esperado de X es

$$\mu = E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

Siempre y cuando la anterior integral exista.

La varianza de la variable aleatoria X se denota y se define así:

$$\sigma^2 = Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$$

Siempre que el valor esperado de X exista.

Para una variable aleatoria discreta X se tiene que

$$\mu = E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

Para una variable aleatoria continua X resulta que

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx$$

Al número,

$$\sigma_X = \sqrt{Var(X)}$$

se le denomina la desviación estándar de la variable aleatoria X .

Expresiones de forma explícita para la varianza de X correspondientes al caso discreto o continuo son, respectivamente,

$$\sigma^2 = Var(X) = E(X - E(X))^2 = \sum_{x \in R_X} (x - \mu)^2 P(X = x)$$

$$\sigma^2 = Var(X) = E(X - E(X))^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)dx$$

El r -ésimo momento central de X alrededor de cero se denota y define así:

$$\mu_r = E(X^r)$$

Siempre y cuando el valor esperado exista.

El r -ésimo momento central de X alrededor de cero para el caso discreto o continuo está dado respectivamente por las expresiones:

$$\mu_r' = E(X^r) = \sum_{x \in R_X} x^r P(X = x)$$

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f_X(x) dx$$

El r -ésimo momento central de X alrededor de $E(X)$ se denota y define así:

$$\mu_r = E((X - E(X))^r)$$

Siempre y cuando el valor esperado exista.

Para el caso discreto y continuo se tienen:

$$\mu_r = E((X - E(X))^r) = \sum_{x \in R_X} (x - \mu)^r P(X = x)$$

$$\mu_r = E((X - E(X))^r) = \int_{-\infty}^{\infty} (x - \mu)^r f_X(x) dx$$

De acuerdo con Abramowitz y Stegun (1972), el coeficiente de asimetría y el coeficiente de curtosis se definen mediante las siguientes expresiones, respectivamente:

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} = E(X - \mu)^3 / \sigma^3$$

$$\alpha_4 = \frac{\mu_4}{(\mu_2)^2} = E(X - \mu)^4 / \sigma^4$$

2.7 Métodos de simulación de variables aleatorias

Enseguida se presentan de manera teórica y con algún nivel de detalle los principales métodos utilizados para simular valores de variables aleatorias, tales como: el método de la transformada inversa, el de rechazo, el de Box Muller, composición y convolución.

2.7.1 Método de la transformada inversa

Si U es una variable aleatoria uniforme en el intervalo $(0,1)$, entonces para cualquier función de distribución F invertible, la variable aleatoria X definida por medio de la expresión:

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

tiene la misma distribución F (Ross, 1999).

En este contexto, para un valor u de la variable aleatoria U , $F^{-1}(u)$ se define como el valor x tal que $u = F(x)$; esto quiere decir que es posible establecer una correspondencia de la siguiente forma:

$$u = F(x) = \int_{-\infty}^x f(t)dt$$

Debido a que $F(x)$ es una función monótona creciente, entonces resulta que

$$F^{-1}(u) = F^{-1}(F(x)) = x \quad (2.1)$$

La expresión (2.1) indica que x es un valor de la variable aleatoria X con función de distribución $F(x)$.

En efecto, sea F_x la función de distribución de $X = F^{-1}(U)$, entonces,

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

Como F es una función de distribución, se tiene que $F(x)$ es una función monótona creciente de x , y , por lo tanto, la desigualdad $a \leq b$ es equivalente a la desigualdad $F(a) \leq F(b)$, luego se puede escribir:

$$F_x(x) = P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x)$$

2.7.2 Método del rechazo

En concordancia con Ross (1999), se supone que ya se tiene un método para generar los valores de una variable aleatoria con una función de densidad de probabilidad $g(x)$; con base en ella se pueden generar los valores x de una variable aleatoria X con función de densidad continua $f(x)$ aceptando el valor con una probabilidad proporcional a $\frac{f(Y)}{g(Y)}$.

Para posibilitar la generación de un algoritmo, se procede de la siguiente manera: sea k una constante que satisface la desigualdad que se indica a continuación:

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

El algoritmo se conforma de los siguientes pasos:

1. Se genera un valor y de la variable aleatoria Y con densidad g .
2. Se genera un valor u de una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$.
3. Si se cumple la desigualdad:

$$U \leq \frac{f(Y)}{kg(Y)} \text{ entonces hacer } X = Y.$$

De lo contrario, regresar al paso 1.

A continuación se prueba que la variable aleatoria X generada por el método del rechazo tiene función de densidad f .

Sea N el número de iteraciones para obtener el valor x de una variable aleatoria X , entonces,

$$P(X \leq x) = P(Y_N \leq x) = P\left(Y \leq x / U \leq \frac{f(Y)}{kg(Y)}\right)$$

De la definición de probabilidad condicional resulta:

$$P(X \leq x) = \frac{P\left(Y \leq x, U \leq \frac{f(Y)}{kg(Y)}\right)}{P\left(U \leq \frac{f(Y)}{kg(Y)}\right)}$$

La anterior expresión se puede escribir de la siguiente forma:

$$P(X \leq x) = \frac{P\left(Y \leq x, U \leq \frac{f(Y)}{kg(Y)}\right)}{h}$$

Donde,

$$h = P\left(U \leq \frac{f(Y)}{kg(Y)}\right)$$

Como las variables aleatorias Y con U son independientes, entonces su función de densidad conjunta cumple que

$$f(y, u) = g(y) \text{ con } 0 < u < 1$$

Con esto se tiene que

$$P(X \leq x) = \frac{1}{h} \iint_{\substack{y \leq x \\ 0 \leq u \leq \frac{f(y)}{kg(y)}}} g(y) du dy$$

La anterior expresión se puede expresar de la siguiente manera:

$$P(X \leq x) = \frac{1}{h} \int_{-\infty}^x \left(\int_0^{\frac{f(y)}{kg(y)}} du \right) g(y) dy$$

Evaluando la integral del paréntesis, resulta

$$P(X \leq x) = \frac{1}{hk} \int_{-\infty}^x f(y) dy$$

Usando el hecho de que f es una función de densidad, haciendo que x tienda al infinito, se obtiene que

$$1 = \frac{1}{hk} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{hk} \Rightarrow hk = 1$$

Al reemplazar en la anterior, finalmente se tiene que

$$P(X \leq x) = \int_{-\infty}^x f(y) dy$$

La anterior, indica X que tiene función de densidad f .

2.7.3 Método de Box Muller

Recogiendo las ideas de Papoulis (1991), este método resulta de utilizar un método basado en coordenadas polares para generar variables aleatorias con distribución normal estándar $N(0,1)$. A continuación se describe la obtención de un algoritmo para este caso.

Sean X y Y variables aleatorias independientes, cada una con distribución normal estándar, como se indican enseguida:

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

Con las anteriores variables se forma un vector aleatorio (X, Y) sobre el plano cartesiano; este mismo vector se puede expresar en las coordenadas polares (R, θ) ; estas coordenadas satisfacen las siguientes igualdades:

$$\begin{aligned} X &= R \cos \theta \\ Y &= R \sin \theta \end{aligned}$$

De las anteriores igualdades se tiene que

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

$$\theta = \tan^{-1}\left(\frac{Y}{X}\right)$$

Sean X y Y variables aleatorias independientes; la función de densidad conjunta es el producto de sus correspondientes funciones de densidad, es decir,

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

Es decir,

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

La anterior expresión se puede escribir como

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

Con el propósito de obtener la función de densidad conjunta de las variables aleatoria R^2 y θ , la cual se denotará con $f(r, \theta)$, se hace el siguiente cambio de variables:

$$r = x^2 + y^2$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Al calcular el jacobiano de la anterior transformación se tiene:

$$J = \begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} = \begin{vmatrix} 2x & 2y \\ -\frac{y}{x^2} \cos^2 \theta & \frac{\cos^2 \theta}{x} \end{vmatrix} = \frac{2x \cos^2 \theta}{x} + \frac{2y^2}{x^2} \cos^2 \theta$$

$$J = 2 \cos^2 \theta + 2 \left(\frac{y}{x}\right)^2 \cos^2 \theta = 2 \cos^2 \theta + 2 \tan^2 \theta \cos^2 \theta = 2$$

Luego la función de densidad conjunta de las variables aleatorias R^2 y θ es

$$f(r, \theta) = \frac{1}{2} \frac{1}{2\pi} \exp\left(-\frac{1}{2}r\right) \quad \text{con } 0 < r < \infty, \quad 0 < \theta < 2\pi$$

La función anterior se puede escribir como el producto de sus densidades marginales por la independencia de las variables R^2 y θ :

$$f(r, \theta) = f(r)f(\theta) = \left(\frac{1}{2}\exp\left(-\frac{1}{2}r\right)\right)\left(\frac{1}{2\pi - 0}\right)$$

$$\text{con } 0 < r < \infty, 0 < \theta < 2\pi$$

De aquí se tiene que

$$f(r) = \frac{1}{2}\exp\left(-\frac{1}{2}r\right) = \lambda \exp(-\lambda r) \quad \text{con } 0 < r < \infty,$$

Así, la variable R^2 sigue una distribución exponencial con parámetro igual a 2.

Su función de distribución es:

$$F(r) = 1 - \exp\left(-\frac{1}{2}r\right) \quad \text{con } 0 < r < \infty,$$

Además,

$$f(\theta) = \frac{1}{2\pi - 0} \quad \text{con } 0 < \theta < 2\pi$$

indica que la variable θ sigue una distribución uniforme en el intervalo $(0, 2\pi)$ de media igual a π . Su función de distribución es:

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

Ahora, si U_1 y U_2 son dos variables aleatorias independientes, cada una con distribución uniforme en el intervalo $(0,1)$, entonces se pueden generar las variables aleatorias R^2 y θ mediante las expresiones

$$\begin{aligned} R^2 &= -2\text{Ln}U_1 \\ \theta &= 2\pi U_2 \end{aligned}$$

Luego para generar las variables aleatorias independientes X y Y con distribución normal estándar se puede optar por aplicar el siguiente algoritmo:

1. Generar U_1 y U_2
2. Hacer:

$$\begin{aligned} R^2 &= -2\text{Ln}U_1 \\ \theta &= 2\pi U_2 \end{aligned}$$

3. Transformar las anteriores variables nuevamente a coordenadas cartesianas (rectangulares), así:

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

$$Y = R \operatorname{sen} \theta = \sqrt{-2 \ln U_1} \operatorname{sen}(2\pi U_2)$$

Las anteriores expresiones conforman el denominado método de Box-Muller.

Algunas veces el método de Box-Muller presenta poca eficiencia computacional; en este caso es recomendable utilizar una forma alternativa propuesta por Marsaglia, que consiste en lo siguiente:

Si U es una variable aleatoria con distribución uniforme en $(0,1)$, entonces la variable $2U$ tiene distribución uniforme en el intervalo $(0, 2)$, y la variable $2U-1$ tiene distribución uniforme en el intervalo $(-1,1)$; al generar valores de dos variables aleatorias U_1 y U_2 con distribución uniforme en el intervalo $(0,1)$ se pueden conformar las siguientes variables aleatorias:

$$V_1 = 2U_1 - 1$$

$$V_2 = 2U_2 - 1$$

El par (V_1, V_2) tiene distribución uniforme en el cuadrado de lado igual a 2 con sus vértices en los puntos $(-1,-1)$, $(-1,1)$, $(1,1)$ y $(1,-1)$. Se generan continuamente pares (V_1, V_2) hasta que se cumpla la condición de que $T = V_1^2 + V_2^2 \leq 1$, es decir, hasta que T resulte menor o igual que 1. Luego, se calcula

$$W = \sqrt{\frac{-2 \ln T}{T}}$$

Finalmente, las variables X, Y con distribución normal estándar se obtienen mediante las expresiones siguientes:

$$X = WV_1$$

$$Y = WV_2$$

Lo anterior permite construir el siguiente algoritmo:

1. Se generan U_1 y U_2 con distribución uniforme en el intervalo $(0,1)$
2. Calcular $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $T = V_1^2 + V_2^2$
3. Si $T > 1$, regresar al paso 1; en caso contrario,
4. Calcular $W = \sqrt{\frac{-2 \ln T}{T}}$
5. Calcular $X = WV_1$, $Y = WV_2$

2.7.4 Método de la composición

Al *método de composición*, también se le denomina *método de mezclas*; para aplicarlo se tiene en cuenta lo siguiente:

Si $F_1(x), F_2(x), F_3(x), \dots, F_n(x)$ son funciones de distribución, y la variable aleatoria de interés X tiene función de distribución $F(x)$, la cual puede expresarse de la siguiente forma:

$$F(x) = c_1 F_1(x) + c_2 F_2(x) + c_3 F_3(x) + \dots + c_n F_n(x)$$

Donde $c_1 + c_2 + c_3 + \dots + c_n = 1$, $c_i > 0$, entonces $F(x)$ es una composición o mixtura (mezcla) de las funciones de distribución $F_i(x)$. Si $f(x)$ es la función de densidad de probabilidad correspondiente a $F(x)$, entonces también es posible tener la siguiente igualdad:

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x) + \dots + p_n f_n(x)$$

Con $p_1 + p_2 + p_3 + \dots + p_n = 1$, $p_i > 0$, a los valores anteriores no negativos cuya suma es uno se les denomina pesos para las funciones de densidad, las $f_i(x)$ son las densidades correspondientes a las funciones de distribución $F_i(x)$. Lo anterior indica que el área bajo la curva $f(x)$ se puede dividir en una cantidad finita de áreas cuya suma debe ser igual a 1.

Un algoritmo para realizar el proceso de simulación es el siguiente:

1. Identificar las subáreas correspondientes a la función de densidad $f(x)$ original.
2. Determinar la función de densidad $f_i(x)$ para cada subárea.
3. Conformar

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x) + \dots + p_n f_n(x)$$

4. Obtener la función de distribución para las subáreas,

$$F(x) = c_1 F_1(x) + c_2 F_2(x) + c_3 F_3(x) + \dots + c_n F_n(x)$$

5. Generar dos números aleatorios U_1, U_2 con distribución uniforme en $(0,1)$.
6. Usar el número aleatorio U_1 para escoger la función de densidad $f_i(x)$ a partir de la cual se simulará un valor para la variable aleatoria X siguiendo el criterio

$$\begin{aligned}
0 < U_1 \leq p_1 &\Rightarrow \text{Generar } x \text{ a partir de } F_1(x) \\
p_1 < U_1 \leq p_1 + p_2 &\Rightarrow \text{Generar } x \text{ a partir de } F_2(x) \\
p_1 + p_2 < U_1 \leq p_1 + p_2 + p_3 &\Rightarrow \text{Generar } x \text{ a partir de } F_3(x) \\
&\vdots \\
&\vdots \\
p_1 + p_2 + p_3 + \dots + p_{n-1} < U_1 \leq p_1 + p_2 + \dots + p_n &\Rightarrow \text{Generar } x \text{ a partir de } F_n(x)
\end{aligned}$$

7. Utilizar el número aleatorio U_2 para simular un valor de X por medio de la distribución $F_i(x)$ correspondiente a $f_i(x)$; para ello se puede utilizar el método de la transformada inversa u otro que se considere adecuado.

Por otro lado, en concordancia con Ross (1999), una manera de simular los valores de una variable aleatoria X a partir de $F(x)$ consiste en simular una variable aleatoria I , igual a i con probabilidad c_i , $i = 1, 2, \dots, n$, y luego simular a partir de $F_i(x)$ (es decir, si el valor simulado de I es $I = j$, entonces la segunda simulación es a partir de $F_j(x)$).

2.7.5 Método de convolución

Este método se aplica cuando una variable aleatoria X se puede expresar como una suma lineal de variables aleatorias, es decir, X se expresa como una combinación lineal de la siguiente manera:

$$X = c_1 X_1 + c_2 X_2 + c_3 X_3 + \dots + c_n X_n$$

Para simular un valor de la variable aleatoria X se requiere generar n valores, que se denotan con $u_1, u_2, u_3, \dots, u_n$, de una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$, para con ellos generar los valores correspondientes a las variables $X_1, X_2, X_3, \dots, X_n$ y, finalmente, obtener el valor de X . En el capítulo tres se ilustran por medio de ejemplos numéricos los métodos descritos anteriormente.

2.8 Método de simulación Monte Carlo

Siguiendo las ideas expresadas por Gentle (1998) y por Ross (1999), una de las primeras aplicaciones de los números aleatorios fue el cálculo de integrales definidas en el intervalo $(0,1)$. Se utiliza el concepto de muestra aleatoria y la ley de los grandes números para construir un algoritmo que proporciona una solución aproximada al valor de una integral definida. El algoritmo resulta interesante especialmente en el cálculo de aquellas integrales que por métodos de cuadratura (analíticos) no se pueden resolver; sin embargo, también el algoritmo ayuda a obtener soluciones con un buen nivel de aproximación para integrales definidas en el intervalo $(0, 1)$, así como en el intervalo (a, b) , usando números aleatorios.

2.8.1 Concepto de muestra aleatoria

En concordancia con Mayorga (2003), una muestra aleatoria proveniente de una población con distribución $F(x)$ es una sucesión de variables aleatorias X_1, X_2, \dots, X_n independientes e idénticamente distribuidas $F(x)$. Esto quiere decir que cada variable aleatoria tiene la misma distribución de probabilidad y son independientes como funciones medibles.

2.8.2 Ley fuerte de los grandes números

Diversos autores mencionan que la *ley fuerte de los grandes números* es la ley más conocida de la teoría de la probabilidad. Según Blanco (2004), esta ley establece que el promedio de una sucesión de variables aleatorias X_1, X_2, \dots, X_n ...independientes e idénticamente distribuidas $F(x)$ con media finita μ y varianza finita σ^2 , converge con probabilidad 1 a la media de la distribución. Lo anterior se escribe de la siguiente manera:

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \sum_{i=1}^n \frac{X_i}{n} \rightarrow \mu \quad \text{cuando } n \rightarrow \infty$$

2.8.3 Un algoritmo para la simulación Monte Carlo

Dada una función $f(x)$ continua en $(0,1)$, se requiere calcular la integral I que se indica a continuación:

$$I = \int_0^1 f(x) dx$$

Para calcular el valor de I se considera una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$. Luego I se puede expresar como el siguiente valor esperado (Ross, 1999):

$$I = E(f(U))$$

Si U_1, U_2, \dots, U_n es una sucesión de variables aleatorias que conforman una muestra aleatoria proveniente de la distribución uniforme en $(0,1)$, entonces $f(U_1), f(U_2), \dots, f(U_n)$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media igual a I . En consecuencia, por la ley fuerte de los grandes números se tiene con probabilidad 1 que

$$\sum_{i=1}^n \frac{f(U_i)}{n} \rightarrow E(f(U)) = I \quad \text{cuando } n \rightarrow \infty \quad (22)$$

La expresión (2.2) posibilita obtener aproximaciones de la integral I al generar gran cantidad de números aleatorios u_i y tomar como una aproximación de I al promedio

de las $f(u_i)$. Este método para aproximar integrales definidas en el intervalo $(0,1)$ se conoce como *de Monte Carlo* (Gentle, 1998; Ross, 1999).

El anterior contexto permite construir el siguiente algoritmo para calcular integrales en $(0,1)$ de manera aproximada, utilizando números aleatorios:

1. Especificar $f(x)$
2. Indicar n
3. Inicializar Suma = 0
4. Para $i = 1$ hasta n hacer
 - Generar u_i
 - Suma = Suma + $f(u_i)$
 - Imprimir $I = \text{Suma}/n$
5. Terminar

2.8.4 Aproximación de una integral en el intervalo $(0,1)$

Se requiere calcular la integral

$$I = \int_0^1 10e^{-x^2} dx$$

El integrando de I es la función continua en $(0,1)$ siguiente:

$$f(x) = 10e^{-x^2}$$

Con los números aleatorios generados y presentados en la Tabla 4.5, el algoritmo para aproximar la integral I es:

1. Hacer $n=50$
Suma=0
2. Para $i=1$ hasta 50 hacer:
 - Generar u_i
 - Suma=Suma + $f(u_i)$
 - Imprimir $I = \text{Suma}/50 = 381.7179/50 = 7.634358$
3. Terminar

Los números aleatorios u_i y los valores $f(u_i)$ se presentan en la Tabla 2.2.

En el algoritmo anterior

$$f(u_i) = 10e^{-u_i^2}$$

se tiene como resultado la aproximación

$$I = \int_0^1 10e^{-x^2} dx \cong 7.634358$$

El cálculo de la integral I , definida en el intervalo $(0,1)$ también se efectuó usando el mismo algoritmo, pero con 100 números aleatorios de una distribución uniforme en $(0,1)$ cuyo intervalo de confianza del 95% fue $(0.423, 0.543)$, y se obtuvo un valor aproximado de 7.533. Se observa que a medida que se aumenta la cantidad de números aleatorios se mejora en la aproximación del cálculo de la integral a su verdadero valor. De manera automática, usando 500 números aleatorios provenientes de una distribución uniforme en $(0,1)$ para los cuales el intervalo de confianza del 95% fue $(0.465, 0.517)$, se obtuvo un valor de 7.50384.

Tabla 2.2. Números aleatorios y su valor en la función $f(u) = 10e^{-u^2}$

N.º	u_i	$f(u_i)$	N.º	u_i	$f(u_i)$
1	0,559355	7,313388	26	0,761206	5,602153
2	0,982616	3,8078	27	0,803134	5,246496
3	0,325262	8,99609	28	0,702858	6,101751
4	0,535503	7,506895	29	0,047913	9,97707
5	0,507678	7,727988	30	0,394261	8,560371
6	0,281995	9,235583	31	0,22862	9,490752
7	0,347206	8,864311	32	0,726215	5,90144
8	0,921085	4,281006	33	0,882465	4,589818
9	0,408148	8,46551	34	0,182222	9,673404
10	0,095739	9,90876	35	0,27969	9,247551
11	0,333891	8,945064	36	0,695545	6,164468
12	0,620771	6,80208	37	0,902572	4,428007
13	0,641131	6,629543	38	0,122813	9,8503
14	0,752135	5,67958	39	0,507697	7,727838
15	0,204952	9,588646	40	0,117283	9,863388
16	0,143259	9,796859	41	0,924725	4,252337
17	0,445881	8,197059	42	0,517283	7,652278
18	0,369064	8,726611	43	0,271406	9,289865
19	0,48528	7,901785	44	0,644636	6,599733
20	0,099268	9,901943	45	0,643786	6,60697
21	0,725651	5,906279	46	0,852829	4,832039
22	0,037251	9,986133	47	0,06515	9,957644
23	0,174697	9,699419	48	0,406798	8,474829
24	0,288352	9,202162	49	0,566701	7,253145
25	0,66887	6,39296	50	0,84329	4,910849

(Los valores de esta tabla fueron generados por los autores con el software libre R).

2.8.5 Método de Monte Carlo para integrales definidas (a,b)

El método de Monte Carlo se puede extender al cálculo aproximado de integrales definidas en el intervalo (a,b), es decir,

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

En este caso, se puede realizar la sustitución: $t = \frac{x-a}{b-a}$, de donde resulta que $(b-a)dt = dx$ con $t \in (0,1)$

$$\text{Luego} \quad I = \int_a^b f(x)dx = \int_0^1 (b-a)f((b-a)t+a)dt \quad (2.3)$$

2.8.6 Aproximación de una integral el intervalo (-2,2)

Se necesita calcular de manera aproximada la integral definida por

$$I = \int_{-2}^2 0.4e^{-0.5x^2} dx$$

La función que se debe integrar es

$$f(x) = 0.4e^{-0.5x^2}$$

La sustitución por realizar está dada por

$$t = \frac{x-(-2)}{2-(-2)} = \frac{x+2}{4}$$

Desarrollando la expresión (2.3) resulta:

$$\begin{aligned} I &= \int_{-3}^2 f(x)dx = \int_0^1 (2-(-2))f((2-(-2))t+(-2))dt \\ I &= \int_{-3}^2 f(x)dx = \int_0^1 4(0.4)f(4t-2)dt = \int_0^1 1.6e^{-0.5(4t-2)^2} dt \end{aligned}$$

Usando los números aleatorios de la Tabla 4.5, el algoritmo para aproximar la integral I es

1. Hacer $n=50$
Suma=0
2. Para $i=1$ hasta 50 hacer:
Generar u_i
Suma=Suma + $1.6e^{-0.5(4u_i-2)^2}$
Imprimir $I = \text{Suma}/50 = 50.24/50 = 1.0048$
3. Terminar

Los números aleatorios u_i y los valores $f(u_i)$ se indican en la Tabla 2.3. Así,

$$I = \int_{-2}^2 0.4e^{-0.5x^2} dx \cong 1.0048$$

En la lectura de la Tabla 2.3 se debe tener en cuenta que

$$f(u_i) = 1.6e^{-0.5(4u_i-2)^2}$$

El cálculo de la integral I definida en el intervalo $(-2,2)$ también se realizó utilizando el mismo algoritmo, pero con 100 números aleatorios de una distribución uniforme en $(0,1)$ cuyo intervalo de confianza del 95% fue $(0.453, 0.565)$, obteniéndose un valor aproximado de 0.9854.

Tabla 2.3. Números aleatorios y su valor en la función $f(x)$ en $(-2,2)$

N°.	u_i	$f(u_i)$	N°.	u_i	$f(u_i)$
1	0,559355	1,555534	26	0,761206	0,92698
2	0,982616	0,248245	27	0,803134	0,767114
3	0,325262	1,253244	28	0,702858	1,151186
4	0,535503	1,583947	29	0,047913	0,311904
5	0,507678	1,599246	30	0,394261	1,4631
6	0,281995	1,09395	31	0,22862	0,887655
7	0,347206	1,327417	32	0,726215	1,062491
8	0,921085	0,387324	33	0,882465	0,49647
9	0,408148	1,495573	34	0,182222	0,713295
10	0,095739	0,432828	35	0,27969	1,085142
11	0,333891	1,283084	36	0,695545	1,178332
12	0,620771	1,423784	37	0,902572	0,437575
13	0,641131	1,364325	38	0,122813	0,512654
14	0,752135	0,96216	39	0,507697	1,599242
15	0,204952	0,797382	40	0,117283	0,495705
16	0,143259	0,578043	41	0,924725	0,3779
17	0,445881	1,562946	42	0,517283	1,596181
18	0,369064	1,394937	43	0,271406	1,053335
19	0,48528	1,597229	44	0,644636	1,353436
20	0,099268	0,442777	45	0,643786	1,356095
21	0,725651	1,064662	46	0,852829	0,591023
22	0,037251	0,288492	47	0,06515	0,352484
23	0,174697	0,68621	48	0,406798	1,492587
24	0,288352	1,118114	49	0,566701	1,544054
25	0,66887	1,273623	50	0,84329	0,623266

(Los valores de esta tabla fueron generados por los autores con el software libre R).

Ejercicios capítulo 2

- 2.1 Proporcione dos ejemplos de experimento aleatorio, espacio muestral, sigma álgebra, medida de probabilidad y espacio de probabilidad.
- 2.2 Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad sobre Ω , probar que:
- Para todo $T \in \mathfrak{F}$ se cumple que $T(T^c) = 1 - P(T)$
 - $P(\emptyset) = 0$
 - Para todo $T, S \in \mathfrak{F}$, $P(T \cup S) = P(S) - P(T \cap S)$
- 2.3 Proporcione un ejemplo de variable aleatoria discreta. Construya la función de probabilidad y la distribución de probabilidad, obtenga su valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- 2.4 Proporcione un ejemplo de variable aleatoria continua. Construya la función de probabilidad y la distribución de probabilidad; obtenga su valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- 2.5 Para una variable aleatoria X con función de probabilidad dada por

Tabla 2.4. Función de probabilidad

x	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{2}{8}$

- Calcular $E(X)$, $Var(X)$ y su desviación estándar
 - Determinar la función de distribución $F_X(x)$
 - Obtener en coeficiente de asimetría y curtosis
- 2.6. Sea X una variable aleatoria, una función definida por
- $$f(x) = \begin{cases} x & \text{si } 0 < x < 1 \\ 2 - x & \text{si } 1 \leq x < 2 \\ 0 & \text{en otro caso} \end{cases}$$
- Verificar si corresponde a una función de densidad.
 - De ser así, obtener la función de distribución de la variable aleatoria X .
 - Obtener el valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- 2.7 Usando el método de Monte Carlo, calcular la integral de la función que se presenta en el ejercicio 2.6 en el intervalo $(-1,5)$.

3

SIMULACIÓN CON VARIABLES ALEATORIAS PROVENIENTES DE DISTRIBUCIONES DISCRETAS Y CONTINUAS

En este capítulo se describen algunos modelos de probabilidad, discretos y continuos, para la simulación de valores de variables aleatorias que se ajustan a ellos. De manera integrada, se diseñan los algoritmos correspondientes y se generan valores aleatorios en cada caso (algunas veces de forma manual, y otras con la ayuda de programas de computador); estos se presentan en tablas construidas para tal fin. En el proceso de simulación se utilizan los métodos descritos en el capítulo anterior.

3.1 Uso de modelos aleatorios discretos en simulación

3.1.1 Modelo de Bernoulli

Una variable aleatoria X tiene la distribución de Bernoulli si su función de probabilidad es

$$f(x, p) = p^x (1 - p)^{1-x}$$

Con:

$$x = 0, 1$$

p : probabilidad de éxito

$1 - p$: probabilidad de fracaso

Su valor esperado y su varianza están dados por:

$$E(X) = p$$
$$Var(X) = p(1 - p)$$

Para generar los valores x de una variable aleatoria Bernoulli X con probabilidad de éxito p , se puede proceder a través de un algoritmo conformado por los siguientes pasos (Barceló, 1996):

1. Generar r_i
2. Asignar, $x_i = 0$ si $r_i > p$
 $x_i = 1$ si $r_i \leq p$

Donde los r_i son números aleatorios con distribución uniforme en el intervalo (0,1). Como ejemplo, en la Tabla 3.1 se presentan los resultados obtenidos al desarrollar un proceso para simular 10 valores x_i de una variable X con distribución de Bernoulli con parámetro $p = 0,7$. La función de probabilidad correspondiente está dada por:

$$f(x, 0.7) = (0.7)^x (0.3)^{1-x}$$

El valor esperado y la varianza son, respectivamente,

$$E(X) = 0.7$$

$$Var(X) = 0.7(0.3) = 0.21$$

Tabla 3.1. Valores simulados para una variable Bernoulli con $p=0.7$

Números aleatorios	P	x_i
0,914	$> 0,7$	0
0,775	$> 0,7$	0
0,495	$< 0,7$	1
0,534	$< 0,7$	1
0,585	$< 0,7$	1
0,185	$< 0,7$	1
0,521	$< 0,7$	1
0,079	$< 0,7$	1
0,968	$> 0,7$	0
0,346	$< 0,7$	1

(Los valores de esta tabla fueron generados por los autores con el software libre R).

Siguiendo un procedimiento similar, en la Tabla 3.2 se indican los resultados de simular 10 valores x_i de una variable X con distribución de Bernoulli de parámetro $p = 0.3$.

La función de probabilidad respectiva corresponde a la siguiente expresión:

$$f(x, 0.3) = (0.3)^x (0.7)^{1-x}$$

El valor esperado y la varianza para este caso son:

$$E(X) = 0.3$$

$$Var(X) = 0.3(0.7) = 0.21$$

Tabla 3.2. Valores simulados para una variable Bernoulli con $p=0.3$

Números aleatorios	p	x_i
0,534	> 0,3	0
0,391	> 0,3	0
0,695	> 0,3	0
0,996	> 0,3	0
0,925	> 0,3	0
0,278	< 0,3	1
0,407	> 0,3	0
0,363	> 0,3	0
0,37	> 0,3	0
0,52	> 0,3	0

(Los valores de esta tabla fueron generados por los autores con el software libre R).

3.1.2 Modelo binomial

Una variable aleatoria X tiene la distribución binomial de parámetros n y p , donde p es la probabilidad de éxito si su función de probabilidad es:

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

Con $x = 0, 2, 3, \dots, n$

$$q = 1 - p$$

$$E(X) = np$$

$$Var(X) = npq$$

Para realizar el proceso de simulación se aplica el siguiente algoritmo:

$$x_i = (x_{i-1}) + 1 \quad \text{si } x_i \leq p$$

$$x_i = (x_{i-1}) \quad \text{si } x_i > p$$

$$i = 1, 2, 3, \dots, n; \quad x_0 = 0$$

A continuación se presenta un ejemplo para simular 10 valores x de una variable X con distribución binomial de parámetro $n = 5$, $p = 0,6$. En los datos que se presentan a continuación, el símbolo $>$ indica que el número aleatorio es mayor que 0,6, y, por lo tanto, se asigna cero (0), y el símbolo \leq indica que el número aleatorio es menor o igual que 0,6, y, por consiguiente, se asigna uno (1).

0,602 > 0	0,026 = 1	0,373 = 1	0,841 > 0
0,714 > 0	0,463 = 1	0,367 = 1	0,753 > 0
0,092 = 1	0,927 > 0	0,658 > 0	0,863 > 0
0,268 = 1	0,02 = 1	0,99 > 0	0,132 = 1
0,091 = 1	0,231 = 1	0,232 = 1	0,478 = 1
<hr style="width: 50%; margin: 0 auto;"/> x=3	<hr style="width: 50%; margin: 0 auto;"/> x=4	<hr style="width: 50%; margin: 0 auto;"/> x=3	<hr style="width: 50%; margin: 0 auto;"/> x=2
0,941 > 0	0,047 = 1	0,156 = 1	0,38 = 1
0,871 > 0	0,96 > 0	0,618 > 0	0,138 = 1
0,982 > 0	0,485 = 1	0,958 > 0	0,273 = 1
0,134 = 1	0,42 = 1	0,01 = 1	0,025 = 1
0,984 > 0	0,934 > 0	0,972 > 0	0,381 = 1
<hr style="width: 50%; margin: 0 auto;"/> x=1	<hr style="width: 50%; margin: 0 auto;"/> x=3	<hr style="width: 50%; margin: 0 auto;"/> x=2	<hr style="width: 50%; margin: 0 auto;"/> x=5
0,769 > 0	0,556 = 1		
0,186 = 1	0,31 = 1		
0,367 = 1	0,705 > 0		
0,488 = 1	0,966 > 0		
0,383 = 1	0,409 = 1		
<hr style="width: 50%; margin: 0 auto;"/> x=4	<hr style="width: 50%; margin: 0 auto;"/> x=3		

Como resultado del proceso de simulación se tienen los valores que se indican a continuación:

$$x = 3,4,3,2,1,3,2,5,4,3$$

Para el caso considerado, la función de probabilidad está dada por la expresión

$$f(x, 0.6) = \binom{5}{x} (0.6)^x (0.4)^{5-x}$$

El valor esperado y la varianza son los siguientes:

$$E(X) = 5(0.6) = 3$$

$$Var(X) = 5(0.6)(0.4) = 1.2$$

En Blanco (2004) se describe una manera alternativa para simular los valores x de una variable X con distribución binomial mediante la aplicación del siguiente algoritmo:

1. Entrada:
Parámetros n, p
Número *máximo* de simulaciones.
2. Inicialización:
 $q := 1 - p$
 P es un vector de q^{n+1} con componentes $P(i)$ definidas e inicializadas de la siguiente manera:
 $P(0) := q^n$
3. Para $j = 1$ (incrementando en 1) hasta n hacer:
$$P(j) := \frac{n-j+1}{j} * \frac{p}{q} * P(j-1)$$

 $suma$ es un vector de R^{n+1} con componentes $suma(i)$ definidas e inicializadas de la siguiente manera:
 $suma(0) := P(0)$
4. Para $j = 1$ (incremento en 1) hasta n hacer:
 $suma(j) := suma(j-1) + P(j)$
5. Para $j = 1$ (incrementando en 1) hasta *máximo* hacer:
 $i := 0$
6. Mientras $RND \geq suma(i)$ hacer:
 $i := i + 1$

Escribir i
7. Finalizar.

3.1.3 Modelo de Poisson

Una variable aleatoria X tiene distribución de Poisson de parámetro $\lambda > 0$ si su función de probabilidad es:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Con $x = 0, 2, 3, \dots, n, \dots$

El valor esperado y la varianza para la variable en consideración son:

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

La función de distribución correspondiente a la variable aleatoria X se puede escribir de la siguiente forma:

$$F(x) = \sum_{i=0}^x \frac{\lambda^i e^{-\lambda}}{i!}$$

$$F(x) = e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!}$$

Una forma de realizar el proceso de simulación consiste en formar clases o intervalos para cada valor de i ; luego, se generan números aleatorios y se observa en cuál clase quedan ubicados.

En el siguiente ejemplo, con base en 60 datos correspondientes al número de llegadas a la cola por minuto en un cajero automático, se determinó que dichos datos se ajustaban a una distribución de Poisson con parámetro $\lambda = 0,5$. Se desea simular 40 valores x de esa variable aleatoria X que sigue una distribución de Poisson. La función de probabilidad corresponde a la expresión

$$f(x) = \frac{(0,5)^x e^{-0,5}}{x!}$$

El valor esperado y la varianza para la variable en consideración son:

$$E(X) = 0,5$$

$$Var(X) = 0,5$$

Su función de distribución es:

$$F(x) = e^{-0,5} \sum_{i=0}^x \frac{(0,5)^i}{i!}$$

$$x = 0 \rightarrow F(0) = e^{-0,5} \sum_{i=0}^0 \frac{(0,5)^i}{i!} = 0,6065$$

$$x = 1 \rightarrow F(1) = e^{-0,5} \sum_{i=0}^1 \frac{(0,5)^i}{i!} = 0,6065(1 + 0,5) = 0,9097$$

$$x = 2 \rightarrow F(2) = e^{-0,5} \sum_{i=0}^2 \frac{(0,5)^i}{i!} = 0,6065(1 + 0,5 + 0,125) = 0,9855$$

$$x = 3 \rightarrow F(3) = e^{-0,5} \sum_{i=0}^3 \frac{(0,5)^i}{i!} = 0,6065(1 + 0,5 + 0,125 + 0,208) = 0,9981$$

$$x = 4 \rightarrow F(4) = e^{-0,5} \sum_{i=0}^4 \frac{(0,5)^i}{i!} = 0,9997$$

Con los valores obtenidos se forman los intervalos aleatorios [0.000, 0.6065]; [0.6066, 0.9097]; [0.9098, 0.9855]; [0.9856, 0.9981]; [0.9982, 0.9997] y [0.9998, 1.000]. Se generan 40 números aleatorios con distribución uniforme en (0,1) con buenas propiedades estadísticas y se ubican en el intervalo correspondiente; de acuerdo con el mencionado intervalo se asumen los valores para x , como se observa en la Tabla 3.3.

Los valores simulados resultantes son:

$$x = 0,0,0,1,1,0,0,0,2,0,0,0,0,0,0,0,2,2,0,0,0,0,0,1,0,0,0,0, \\ 1,0,1,0,0,1,2,0,1,3$$

En Blanco (2004) se menciona otra forma de simular los valores x de una variable X con distribución Poisson, la cual consiste en utilizar el siguiente algoritmo:

1. Entrada:

Parámetros λ

Número *máximo* de simulaciones

2. Para $j = 1$ (incrementando en 1) hasta *máximo* hacer:

Inicialización:

$i := 0$

$prob := e^{-\lambda}$

$suma := prob$

3. Verificación

Mientras $RND \geq suma$ hacer:

$i := i + 1$

$prob := \lambda * \frac{prob}{i}$

$suma := suma + prob$

Escribir i

4. Finalizar.

Tabla 3.3. Valores simulados para una variable Poisson de parámetro 0.5

X	Intervalo aleatorio	Números aleatorios generados	Valores variable X
0	0.000-0.6065	0.1160, 0.1000, 0.2830, 0.5238, 0.4132, 0.3206 0.2583, 0.6011, 0.0245, 0.2240, 0.5980, 0.2994 0.4980, 0.0040, 0.5012, 0.3370, 0.2856, 0.4197 0.1230, 0.5480, 0.4902, 0.3869, 0.5674, 0.4246 0.0354, 0.1982, 0.1965, 0.3461	0,0,0,0,0,0 0,0,0,0,0,0 0,0,0,0,0,0 0,0,0,0,0,0 0,0,0,0
1	0.6066-0.9097	0.8525, 0.6714, 0.6219, 0.7637 0.7222, 0.6122, 0.6211	1, 1, 1, 1 1, 1, 1
2	0.9098-0.9855	0.9158, 0.9316, 0.9273, 0.9249	2, 2, 2, 2
3	0.9856-0.9981	0.9962	3
4	0.9982-0.9997		
5	0.9990-1.000		

(Los valores de esta tabla fueron generados por los autores con el software libre R).

3.1.4 Aplicación del método de la transformada inversa para modelos discretos

En este caso se quiere simular valores x de una variable aleatoria X con una distribución dada cuya función de probabilidad se puede expresar de la siguiente forma:

$$P(X = x_i) = p_i \text{ con } i = 0, 1, 2, 3, \dots$$

Además,

$$\sum_i p_i = 1$$

Se supone que el rango R_x de la variable X asume los valores $x_0, x_1, x_2, x_3, \dots$. Para realizar el proceso de simulación se generan valores u correspondientes a una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$; estos se involucran en el proceso como se indica a continuación:

$$X = \begin{cases} x_0 & \text{si } u < p_0 \\ x_1 & \text{si } p_0 \leq u < p_0 + p_1 \\ \bullet \\ \bullet \\ x_i & \text{si } \sum_{j=1}^{i-1} p_j \leq u < \sum_{j=1}^i p_j \\ \bullet \\ \bullet \end{cases}$$

Debido a que la variable aleatoria U tiene distribución uniforme en $(0,1)$, se cumple que

$$P(a \leq U < b) = b - a$$

Con $0 < a < b < 1$. Por la definición de la función de probabilidad resulta que

$$P(X = x_i) = P\left(\sum_{j=1}^{i-1} p_j \leq u < \sum_{j=1}^i p_j\right) = p_i$$

Lo anterior implica que si los x_i están ordenados de manera ascendente, entonces se cumple que

$$X = x_i \text{ siempre y cuando } F(x_{i-1}) \leq u < F(x_i)$$

El contexto anterior permite plantear el siguiente algoritmo para simular los valores x de una variable aleatoria discreta X :

Generar un número aleatorio u

Si $u < p_0$, hacer $x = x_0$ y terminar

Si $u < p_0 + p_1$, hacer $x = x_1$ y terminar; se continúa el proceso

Si $u < p_0 + p_1 + \dots + p_p$, hacer $x = x_i$ y terminar, y así sucesivamente.

A continuación se presenta un ejemplo en el cual se trabaja con el anterior algoritmo. Sea X una variable aleatoria con una distribución cuya función de probabilidad está dada por:

$$f(x) = P(X = x) = \begin{cases} 0.12 & \text{si } x = 1 \\ 0.18 & \text{si } x = 2 \\ 0.30 & \text{si } x = 3 \\ 0.25 & \text{si } x = 4 \\ 0.15 & \text{si } x = 5 \end{cases}$$

Su función de distribución es:

$$F_x(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.12 & \text{si } 1 \leq x < 2 \\ 0.30 & \text{si } 2 \leq x < 3 \\ 0.60 & \text{si } 3 \leq x < 4 \\ 0.85 & \text{si } 4 \leq x < 5 \\ 1 & \text{si } x \geq 5 \end{cases}$$

Los valores x de una variable aleatoria discreta X que se pueden obtener mediante los números aleatorios u de una variable aleatoria discreta U con distribución uniforme en el intervalo $(0,1)$ se pueden obtener mediante la siguiente expresión:

$$x = \begin{cases} 1 & \text{si } u < 0.12 \\ 2 & \text{si } 0.12 \leq u < 0.30 \\ 3 & \text{si } 0.30 \leq u < 0.60 \\ 4 & \text{si } 0.60 \leq u < 0.85 \\ 5 & \text{si } 0.85 \leq u < 1 \end{cases}$$

Con los 10 números aleatorios de la Tabla 3.4 se simularon los 10 valores x para la variable aleatoria en consideración.

Tabla 3.4 Valores simulados de la variable X

Números aleatorios	X
0,564	3
0,106	1
0,685	4
0,096	1
0,988	5
0,45	3
0,757	4
0,213	2
0,729	4
0,367	3

(Los valores de esta tabla fueron generados por los autores con el software libre R).

Adicionalmente, es pertinente aclarar que este método es aplicable a los modelos de Bernoulli, Binomial, Poisson o cualquier modelo discreto (Barceló, 1996).

3.2 Simulación de variables aleatorias continuas

A continuación se indican algunas distribuciones usuales y el método de simulación que frecuentemente resulta más apropiado para simular valores de variables aleatorias que se modelen a través de dichas distribuciones; comúnmente, estas se describen en la mayoría de textos dedicados al tema de la simulación. Luego, a manera de ejemplos se presentan algunas de ellas y los procedimientos correspondientes para realizar la simulación de una cantidad determinada de valores aleatorios.

Para simular valores de variables con distribución uniforme, exponencial, weibull, triangular, empírica, discreta arbitraria o geométrica se utiliza el método de la transformada inversa; para variables con distribución *erlang* o binomial se usa el método de convolución; para variables con distribución gamma, normal, Poisson se puede trabajar con el método del rechazo; la distribución normal también se puede simular mediante transformación directa, las 12 uniformes y el método de Box-Muller.

3.2.1 Modelo uniforme

Una variable aleatoria X sigue la distribución uniforme en el intervalo (a, b) si su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{Si } x \in (a, b) \\ 0 & \text{en otro caso} \end{cases}$$

El valor esperado y la varianza están dados por:

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

En particular, una variable aleatoria X sigue la distribución uniforme en el intervalo $(0,1)$ que se denota con $U(0,1)$, si su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} 1 & \text{Si } x \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$

En este caso, el valor esperado y la varianza toman los valores siguientes:

$$E(X) = \frac{1}{2}$$

$$Var(X) = \frac{1}{12}$$

La función de distribución para la variable aleatoria X que sigue la distribución uniforme en el intervalo (a, b) es:

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

Para simular los valores de una variable aleatoria discreta con distribución uniforme en el intervalo se generan los números aleatorios de una variable aleatoria discreta con distribución uniforme en el intervalo $(0,1)$, y mediante el método de la transformada se aplica la siguiente expresión:

$$u = F(x)$$

O equivalentemente,

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

Luego para $a \leq x < b$ se tiene que

$$u = \frac{x-a}{b-a}$$

De aquí resulta que

$$x = a + u(b-a) = F^{-1}(u)$$

Un algoritmo para simular valores aleatorios de una variable con distribución uniforme en el intervalo (a, b) es el siguiente:

1. Entrada:
 Parámetros a, b
 Número *máximo* de simulaciones
2. Para $j = 1$ (incrementando en 1) hasta *máximo* hacer:
 Generación número aleatorio u
 $x := a + u(b-a)$
3. Terminar.

La función de distribución para la variable aleatoria X que sigue la distribución uniforme en el intervalo $(-2, 2) = (a, b)$ es:

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

A manera de ejemplo, se desea simular 10 valores x de una variable aleatoria discreta X con distribución uniforme en el intervalo $(-2, 2)$. Para ello, se generan 10 números aleatorios u provenientes de una distribución uniforme en el intervalo $(0,1)$ y se reemplazan en la expresión siguiente:

$$x = a + u(b - a) = -2 + u(2 - (-2))$$

De manera simplificada, resulta la expresión

$$x = -2 + 4u$$

Al sustituir los números aleatorios u , resultan los valores simulados que se presentan en la Tabla 3.5.

Tabla 3.5. Valores simulados de una uniforme en $(-2,2)$

Números aleatorios	$x = -2 + 4u$
0,13964	-1,44144
0,4313	-0,2748
0,61218	0,44872
0,29075	-0,837
0,15573	-1,37708
0,6995	0,798
0,3463	-0,6148
0,44564	-0,21744
0,05242	-1,79032
0,1032	-1,5872

(Los valores de esta tabla fueron generados por los autores con el software libre R)

3.2.2 Modelo Exponencial

Una variable aleatoria X tiene distribución exponencial de parámetro $\lambda > 0$ si su función de densidad es:

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

El valor esperado y la varianza de la variable X son:

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

La función de distribución para la variable aleatoria X que sigue la distribución exponencial de parámetro $\lambda > 0$ es:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

Para simular los valores x de una variable aleatoria discreta X con distribución exponencial se generan los números aleatorios u de una variable aleatoria discreta U con distribución uniforme en el intervalo $(0,1)$, y mediante el método de la transformada inversa se aplica la siguiente igualdad:

$$u = F(x)$$

La cual implica la expresión

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

Luego para $\lambda > 0$, $\lambda \geq 0$ se tiene que

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

Despejando x resulta

$$x = \frac{-1}{\lambda} \text{Ln}(1 - u)$$

Un algoritmo para simular valores aleatorios de una variable con distribución exponencial de parámetro es el siguiente:

1. Entrada:
Parámetro $\lambda > 0$
Número *máximo* de simulaciones
2. Para $j = 1$ (incrementando en 1) hasta *máximo* hacer:

Generación número aleatorio u

$$x := \frac{-1}{\lambda} \text{Ln}(1 - u)$$

3. Terminar.

Se requiere simular 10 valores x de una variable aleatoria discreta X con distribución exponencial de parámetro $\lambda = 2$.

En este caso, la función de densidad de probabilidad es:

$$f(x) = \begin{cases} 2e^{-2x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

El valor esperado y la varianza de la variable X son:

$$E(X) = \frac{1}{2}$$

$$Var(X) = \frac{1}{2^2} = \frac{1}{4}$$

La función de distribución para la variable aleatoria X es:

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

Para simular los 10 valores x de una variable aleatoria discreta X en cuestión, se generan 10 números aleatorios u provenientes de una distribución uniforme en el intervalo $(0,1)$ y se reemplazan en la expresión siguiente:

$$x = \frac{-1}{\lambda} \text{Ln}(1-u) = \frac{-1}{2} \text{Ln}(1-u)$$

De forma más simple, se tiene la expresión

$$x = -0.5 \text{Ln}(1-u)$$

Al reemplazar los números aleatorios u resultan los valores simulados que se presentan en la Tabla 3.6.

Observación. La expresión

$$x = \frac{-1}{\lambda} \text{Ln}(1-u)$$

es equivalente a la expresión

$$x = \frac{-1}{\lambda} \text{Ln}(u)$$

debido a que si u toma valores en el intervalo $(0,1)$, entonces también $1-u$ toma valores en el intervalo $(0,1)$. Esta última expresión es también utilizada para simular valores de una variable aleatoria con distribución exponencial de parámetro λ .

Tabla 3.6. Valores simulados de una exponencial de parámetro 2

Números aleatorios	$x = -0.5\text{Ln}(1-u)$
0,486	0,3327
0,187	0,1035
0,317	0,1906
0,295	0,1747
0,788	0,7755
0,129	0,069
0,982	2,0086
0,597	0,4544
0,742	0,6773
0,068	0,0352

(Los valores de esta tabla fueron generados por los autores con el software libre R).

3.2.3 Modelo continuo no usual

Los valores de una variable aleatoria X se ajustan a una función definida de la siguiente manera:

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

La representación de la anterior función se presenta en la Figura 3.1, en la cual se observa una función continua a trozos; con base en esta función, se requiere simular 10 valores de la variable aleatoria X . Para esto, inicialmente se debe mostrar que $f(x)$ es una función de densidad de probabilidad para X .

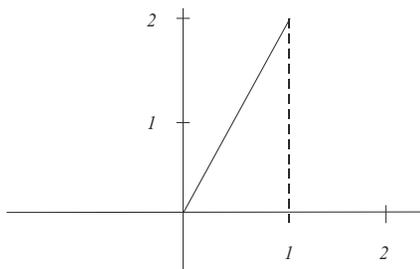


Figura 3.1. Representación gráfica de $f(x)$

Fuente: Los autores, con la ayuda del software libre R.

1. De la Figura 3.1 se observa que $f(x) \geq 0$

$$2. \quad \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^1 f(x)dx + \int_1^{\infty} f(x)dx$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^0 (0)dx + \int_0^1 2x dx + \int_1^{\infty} (0)dx = x^2 \Big|_0^1 = 1^2 - 0^2 = 1$$

De las partes 1. y 2. se tiene que $f(x)$ sí corresponde a una función de densidad de probabilidad para la variable aleatoria X .

Utilizando la expresión

$$u = F(x)$$

resulta

$$u = F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 f(t)dt + \int_0^x f(t)dt$$

$$u = F(x) = \int_0^x 2t dt = t^2 \Big|_0^x = x^2$$

$$u = x^2$$

Para $0 \leq x \leq 1$ se tiene que

$$x = \sqrt{u}$$

Ahora se generan 10 números aleatorios u provenientes de una distribución uniforme en el intervalo $(0,1)$ y se reemplazan en la expresión anterior; los resultados se presentan en la Tabla 3.7.

Tabla 3.7. Valores simulados de un modelo no usual

Números aleatorios	$x = \sqrt{u}$
0,141156	0,37571
0,042938	0,20721
0,621706	0,78848
0,153592	0,39191
0,715214	0,8457
0,928301	0,96348
0,578133	0,76035
0,261857	0,51172
0,72483	0,85137
0,037092	0,19259

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

3.2.4 Aplicación del método de composición

Se necesita generar valores de una variable aleatoria X cuya función de densidad se define de la siguiente forma:

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

Por la forma como está definida la función de densidad, hay dos subáreas correspondientes a la $f(x)$ original; cada una equivale a 0.5. Se puede tomar $p_1 = 0.5$ y $p_2 = 0.5$

Las funciones de densidad $f_1(x)$ para cada subárea deben satisfacer:

$$f(x) = 0.5f_1(x) + 0.5f_2(x)$$

Luego se puede tomar

$$f_1(x) = \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_2(x) = \begin{cases} 2(2-x) & \text{si } 1 \leq x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Las correspondientes distribuciones de probabilidad son:

$$F_1(x) = x^2$$

$$F_2(x) = 4x - x^2 - 3$$

Se generan dos números aleatorios U_1, U_2 con distribución uniforme en $(0,1)$, luego, se aplica el siguiente algoritmo:

Si $0 < U_1 \leq 0.5$, entonces $x^2 = U_1 \Rightarrow x = \sqrt{U_1}$

En caso contrario:

$$4x - x^2 - 3 = U_2 \Rightarrow x = 2 - \sqrt{1 - U_2}$$

A continuación se ilustra el proceso para los tres primeros números aleatorios U_1 , y el uso del número aleatorio U_2 , que se indican en la Tabla 3.8.

$$U_1 = 0.559355 \Rightarrow x = 2 - \sqrt{1 - 0.333891} = 1.183845$$

$$U_1 = 0.982616 \Rightarrow x = 2 - \sqrt{1 - 0.620771} = 1.384184$$

$$U_1 = 0.325262 \Rightarrow x = \sqrt{0.641131} = 0.800707$$

Lo anterior es concordante con los valores x en el dominio de definición de la función de densidad $f(x)$ en los intervalos $(0,1)$ y $(1,2)$.

Tabla 3.8. Valores simulados por el método de composición

u_1	u_2	X
0,559355	0,333891	1,183845
0,982616	0,620771	1,384184
0,325262	0,641131	0,800707
0,535503	0,752135	1,50214
0,507678	0,204952	1,108345
0,281995	0,143259	0,378496
0,347206	0,445881	0,667743
0,921085	0,369064	1,205685
0,408148	0,48528	0,69662
0,095739	0,099268	0,315068

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

3.2.5 Simulación con una distribución Beta

Inicialmente se define la distribución Beta de manera general, para luego realizar un ejemplo de simulación de valores de una variable aleatoria con distribución Beta para un caso específico de sus parámetros.

Una variable aleatoria X tiene distribución Beta de parámetros $a > 0$ y $b > 0$ si su función de densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

Donde $B(a, b)$ es la función Beta definida a través de la siguiente expresión:

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

El valor esperado y la varianza de la variable aleatoria X son:

$$E(X) = \frac{a}{a+b}$$

$$Var(X) = \frac{ab}{(a+b+1)(a+b)^2}$$

Como caso particular, se requiere simular valores x de la variable aleatoria X con distribución Beta de parámetros $a > 0$ y $b = 1$.

En este caso, la función de densidad de probabilidad toma la forma:

$$f(x) = \begin{cases} \frac{1}{B(a,1)} x^{a-1} (1-x)^{1-1} & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

Es decir,

$$f(x) = \begin{cases} \frac{1}{B(a,1)} x^{a-1} & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

Donde,

$$B(a,1) = \int_0^1 x^{a-1} dx = \frac{x^a}{a} \Big|_0^1 = \frac{1}{a}$$

Luego la función de densidad se puede escribir así:

$$f(x) = \begin{cases} ax^{a-1} & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

Su función de distribución es:

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^0 0 dt + \int_0^x f(t) dt = \int_0^x at^{a-1} dt = x^a \quad \text{si } 0 < x < 1$$

Esto es,

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^a & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Luego para simular los valores x de la variable aleatoria X con distribución Beta de parámetros $a > 0$ y $b = 1$ resulta pertinente utilizar el método de la transformada inversa. Para esto, se generan valores u de una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$ y se utiliza la siguiente igualdad:

$$u = F(x) = x^a \quad \text{si } 0 \leq x < 1$$

Lo anterior implica que

$$u = x^a \quad \Rightarrow \quad x = F^{-1}(u) \quad \Rightarrow \quad x = u^{\frac{1}{a}}$$

A manera de ilustración, en la Tabla 3.9 se presentan 20 valores x de la variable aleatoria X , considerando el caso en que $a = 2$, $a = 3$, $a = 4$ y $a = 5$, usando 20 números aleatorios provenientes de una distribución uniforme en $(0,1)$.

Tabla 3.9. Valores simulados de una distribución Beta

U	$a=2$	$a=3$	$a=4$	$a=5$
0,559355	0,7479	0,82394	0,864812	0,890302
0,982616	0,99127	0,994171	0,995625	0,996499
0,325262	0,570317	0,687719	0,755194	0,798816
0,535503	0,731781	0,812058	0,855442	0,882577
0,507678	0,712515	0,797743	0,844106	0,873208
0,281995	0,531032	0,655763	0,728719	0,776333
0,347206	0,589242	0,70285	0,767621	0,809315
0,921085	0,959732	0,972971	0,979659	0,983694
0,408148	0,638865	0,741776	0,79929	0,835918
0,095739	0,309417	0,45747	0,556253	0,625486
0,333891	0,577833	0,693748	0,760153	0,80301
0,620771	0,78789	0,853055	0,887632	0,909047
0,641131	0,800707	0,862281	0,894822	0,914933
0,752135	0,867257	0,909422	0,931266	0,944624
0,204952	0,452716	0,589591	0,672842	0,728334
0,143259	0,378496	0,523248	0,61522	0,677992
0,445881	0,667743	0,763964	0,817156	0,850832
0,369064	0,607506	0,7173	0,779427	0,819257
0,48528	0,69662	0,785834	0,834638	0,865363
0,099268	0,315068	0,463024	0,561309	0,630031

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

3.2.6 Aplicación del método del rechazo

Se requiere simular los valores x de una variable aleatoria X con función de densidad continua $f(x)$ definida de la siguiente manera:

$$f(x) = \begin{cases} 12x(1-x)^2 & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

De la expresión anterior se observa que la función de densidad de la variable aleatoria X se anula fuera del intervalo $(0,1)$; luego para aplicar el método del rechazo es conveniente considerar una función $g(x)$ tal que $f(x) \leq g(x)$ para todo $0 < x < 1$. Para cumplir con el anterior requerimiento se ha de considerar una función de densidad $g(x)$, definida de la siguiente forma:

$$g(x) = \begin{cases} 1 & \text{si } 0 < x < 1 \\ 0 & \text{si } x \notin (0,1) \end{cases}$$

La constante que verifique la siguiente desigualdad $\frac{f(x)}{g(x)} \leq c$ se puede determinar buscando el máximo de

$$\frac{f(x)}{g(x)} = 12x(1-x)^2 = 12x - 24x^2 + 12x^3$$

Derivando e igualando a cero se obtiene que

$$12 - 48x + 36x^2 = 0$$

De donde se tiene que $x = 1$ o $x = \frac{1}{3}$; usando el criterio de la segunda derivada se observa que en $x = \frac{1}{3}$ la segunda derivada es menor que cero, garantizando la existencia de un máximo en $x = \frac{1}{3}$, cuyo valor es:

$$12 \left(\frac{1}{3} \right) \left(1 - \frac{1}{3} \right)^2 = \frac{48}{27} = \frac{16}{9} = c$$

Ahora,

$$\frac{f(x)}{g(x)} \leq c = \frac{16}{9}$$

Por lo tanto,

$$\frac{f(x)}{cg(x)} = \frac{12x(1-x)^2}{\left(\frac{16}{9}\right)} = \frac{108x(1-x)^2}{16} = \frac{27x(1-x)^2}{4}$$

En concordancia con lo expuesto en la sección 2.7.2, el algoritmo que se ha de seguir es:

1. Se genera un valor y de la variable aleatoria Y con densidad $g(x)$ que corresponde a una densidad uniforme en el intervalo $(0,1)$.
2. Se genera un valor u de una variable aleatoria U con distribución uniforme en el intervalo $(0,1)$.
3. Si se cumple la desigualdad:

$$u \leq \frac{27y(1-y)^2}{4} \text{ entonces hacer } x = y.$$

De lo contrario, regresar al paso 1.

En la Tabla 3.10, la primera columna contiene 20 valores generados para una variable aleatoria con función de densidad $g(x)$; la segunda columna presenta 20 números aleatorios con distribución uniforme en $(0,1)$ que se utilizan para comparar con los resultados de la expresión indicada en la tercera columna y permiten generar los valores x de una variable aleatoria X con función de densidad continua $f(x)$ que aparecen en la cuarta columna. La letra N que aparece en la cuarta columna indica que no se generó el valor x porque la desigualdad no se cumplió. En resumen, solo se pudieron generar 13 valores x en este caso.

Tabla 3.10. Valores simulados en la aplicación del método del rechazo

y	u	$27y(1-y)^2/4$	$x = y$
0,559355	0,761206	0,73311	0,559355
0,982616	0,803134	0,002004	N
0,325262	0,702858	0,999557	0,325262
0,535503	0,047913	0,779887	0,535503
0,507678	0,394261	0,830597	0,507678
0,281995	0,22862	0,981296	0,281995
0,347206	0,726215	0,998719	0,347206
0,921085	0,882465	0,038719	N
0,408148	0,182222	0,965045	0,408148
0,095739	0,27969	0,528421	N
0,333891	0,695545	0,999998	0,095739
0,620771	0,902572	0,602613	N
0,641131	0,122813	0,557343	0,641131
0,752135	0,507697	0,31191	N
0,204952	0,117283	0,874465	0,204952
0,143259	0,924725	0,709782	N
0,445881	0,517283	0,924121	0,445881
0,369064	0,271406	0,99169	0,369064
0,48528	0,644636	0,867837	0,48528
0,099268	0,643786	0,543631	N

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

Ejercicios capítulo 3

- 3.1 Simular de forma manual 10 valores de una variable aleatoria con distribución de Bernoulli de parámetros $p=0.15$. Usando el computador, simular 200 valores de la mencionada variable.
- 3.2 Simular 50 valores de una variable aleatoria con distribución binomial de parámetros $n = 18$ y $p = 0.72$.
- 3.3 Simular 30 valores de una variable aleatoria con distribución de Poisson de parámetro 0.8. En el computador, simular 500 valores de dicha variable.
- 3.4 Simular 40 valores de una variable aleatoria discreta cuya función de probabilidad se indica en el ejercicio 2.5.
- 3.5 Simular 35 valores de una variable aleatoria continua cuya función de densidad de probabilidad se presenta en el ejercicio 2.6.
- 3.6 Simular 100 valores de una variable aleatoria con distribución uniforme en el intervalo $(-1, 3)$. Mediante el computador, simular 5000 valores de dicha variable.
- 3.7 Simular 25 valores de una variable aleatoria con distribución exponencial de parámetro 10. Usando el computador, simular 1000 valores de la mencionada variable.
- 3.8. Simular 20 valores de una variable aleatoria con distribución Beta de parámetros $a=9$ y $b=1$. A través del computador, simular 500 valores de la mencionada variable aleatoria.
- 3.9 Usando el método del rechazo, obtener 15 valores de una variable aleatoria con función de densidad dada por

$$f(x) = \begin{cases} 30x^2(1-x)^2 & \text{Si } x \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$

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

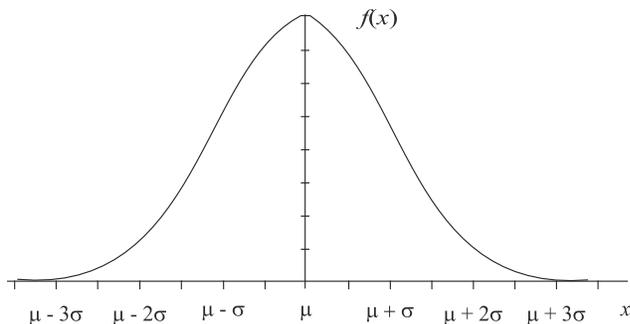


Figura 4.1. Curva normal o campana de Gauss

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

$$y = -\text{Ln}(u)$$

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.

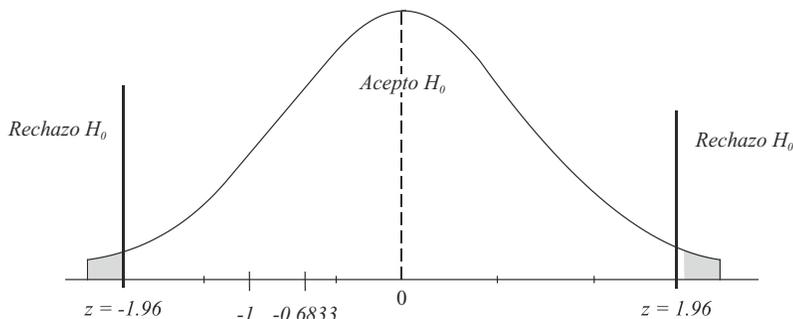


Figura 4.2. Regiones de aceptación y rechazo de H_0

(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.

5

SIMULACIÓN DE VARIABLES ALEATORIAS PROVENIENTES DE LA DISTRIBUCIÓN LAMBDA GENERALIZADA

En este capítulo se hace, inicialmente, una caracterización de la Distribución Lambda Generalizada; para ello se presentan su correspondiente definición y sus cuatro primeros momentos, se proporcionan algunos ejemplos de densidades provenientes de la mencionada distribución y se indican algunas figuras de ellas. Posteriormente, se desarrollan procesos de simulación que permiten generar valores de variables aleatorias con distribución simétrica y asimétrica provenientes de la distribución Lambda Generalizada mediante un algoritmo general basado en el método de la transformada inversa.

5.1 Caracterización de la Distribución Lambda Generalizada

La Distribución Lambda Generalizada (DLG) es una familia de distribuciones muy versátil que permite obtener funciones de densidad con diversas formas, especialmente aquellas de forma acampanada, caracterizadas en muchos casos por ser simétricas, asimétricas, platicúrticas o leptocúrticas, dependiendo de los valores que puedan asumir los cuatro parámetros que las definen o de sus momentos de tercer o cuarto orden, como se indica en las secciones subsiguientes.

5.1.1 Definición de la DLG

Según Karian y Dudewics (2000), la DLG corresponde a una familia de distribuciones de probabilidad, que se define a través de su función percentil como sigue:

$$F^{-1}(y) = x = F^{-1}(y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2} \quad (5.1)$$

Con $0 \leq y \leq 1$, en la definición dada en la expresión (5.1) es necesario el cumplimiento de la condición de que $\lambda_2 \neq 0$. Las constantes $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, se llaman parámetros de la Distribución Lambda Generalizada. El valor λ_1 se asocia con el parámetro de localización; el valor λ_2 , con el parámetro de escala; el valor λ_3 , con el momento de tercer orden, que define la asimetría de una distribución de probabilidad, y el valor λ_4 , con el momento de cuarto orden, que permite definir el apuntamiento de la curva o también denominado curtosis.

De la expresión (5.1) se tiene que $x = F^{-1}(y)$, de donde se deduce que $y = F(x)$.

Si $F(x)$ es la función de distribución de una variable aleatoria continua X , al aplicar cálculo diferencial, derivando con respecto a x mediante el procedimiento de derivación para una función inversa se obtiene la función de densidad siguiente:

$$\frac{d}{dx}(F(x)) = f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \quad ; x = F^{-1}(y)$$

A los cuatro parámetros antes citados se les deben asignar valores específicos apropiados para que las expresiones que definen la Distribución Lambda Generalizada y la función de densidad cumplan con las dos condiciones siguientes que formalmente distinguen una función de densidad de probabilidad:

1. $f(F^{-1}(y)) = f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(F^{-1}(y)) d(F^{-1}(y)) = \int_{-\infty}^{\infty} f(x) dx = 1$

La primera condición indica que la curva $f(x)$ debe estar por arriba del eje horizontal o sobre él, pero nunca por debajo; analíticamente, implica que

$$\frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \geq 0$$

La segunda condición implica que el área bajo la curva es igual a 1.

Para determinar los pares λ_3, λ_4 , en los cuales la Distribución Lambda Generalizada resulte válida, es decir, permita definir formalmente una distribución de probabilidad, Karian y Dudewicz (2000) estudiaron las siguientes regiones:

$$R_1 = \{(\lambda_3, \lambda_4) \ni \lambda_3 \leq -1, \lambda_4 \geq 1\}$$

$$R_2 = \{(\lambda_3, \lambda_4) \ni \lambda_3 \geq 1, \lambda_4 \leq -1\}$$

$$R_3 = \{(\lambda_3, \lambda_4) \ni \lambda_3 \geq 0, \lambda_4 \geq 0\}$$

$$R_4 = \{(\lambda_3, \lambda_4) \ni \lambda_3 \leq 0, \lambda_4 \leq 0\}$$

$$V_1 = \{(\lambda_3, \lambda_4) \ni \lambda_3 < 0, 0 < \lambda_4 < 1\}$$

$$V_2 = \{(\lambda_3, \lambda_4) \ni 0 < \lambda_3 < 1, \lambda_4 < 0\}$$

$$V_3 = \{(\lambda_3, \lambda_4) \ni -1 < \lambda_3 < 0, \lambda_4 > 1\}$$

$$V_4 = \{(\lambda_3, \lambda_4) \ni \lambda_3 > 1, -1 < \lambda_4 < 0\}$$

Karian y Dudewicz demostraron que la Distribución Lambda Generalizada es válida en las regiones: R_1 , R_2 , R_3 , y R_4 , pero no es válida en las regiones V_1 y V_2 ; además, llegaron a la conclusión de que la Distribución Lambda Generalizada es válida en las regiones V_3 y V_4 , si y solo si

$$\frac{(1 - \lambda_3)^{1 - \lambda_3}}{(\lambda_4 - \lambda_3)^{\lambda_4 - \lambda_3}} (\lambda_4 - 1)^{\lambda_4 - 1} < \frac{-\lambda_3}{\lambda_4}$$

El anterior contexto indica que la DLG no está definida para valores arbitrarios de sus cuatro parámetros. Karian y Dudewicz (2000) han propuesto una tabla que contiene una gama muy amplia de valores para los parámetros, de tal manera que con ellos sí se tiene un amplio espectro de distribuciones provenientes de la DLG; así mismo, han desarrollado un algoritmo que permite cubrir una muy amplia variedad de casos de distribuciones que se pueden generar al dar valores específicos a los parámetros de la DLG, resultando distribuciones simétricas, asimétricas, platicúrticas, leptocúrticas u otras que puedan ser útiles en la intención de modelar un determinado conjunto de datos reales.

5.1.2 Momentos de la Distribución Lambda Generalizada

Si la variable aleatoria X tiene Distribución Lambda Generalizada de parámetros $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ con $\lambda_3 > -1/4$ y $\lambda_4 > -1/4$, entonces los cuatro primeros momentos que se denotan con $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ y que determinan la media, la varianza, el sesgo (asimetría) y la curtosis (apuntamiento), respectivamente, se obtienen a través de las siguientes expresiones:

$$\alpha_1 = \mu_X = E(X) = \lambda_1 + \frac{A}{\lambda_2}$$

$$\alpha_2 = \sigma^2 = E[(X - \mu)^2] = \frac{B - A^2}{\lambda_2^2}$$

$$\alpha_3 = E(X - \mu)^3 / \sigma^3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}$$

$$\alpha_4 = E(X - \mu)^4 / \sigma^4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4}$$

Donde, $A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\text{Beta}(1 + \lambda_3, 1 + \lambda_4)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\text{Beta}(1 + 2\lambda_3, 1 + \lambda_4) \\ + 3\text{Beta}(1 + \lambda_3, 1 + 2\lambda_4)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\text{Beta}(1 + 3\lambda_3, 1 + \lambda_4) \\ + 6\text{Beta}(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\text{Beta}(1 + \lambda_3, 1 + 3\lambda_4)$$

Los momentos tercero y cuarto de la distribución permiten determinar la forma de la función de densidad de probabilidad, usando los siguientes criterios:

Si $\alpha_3 = 0$, entonces la distribución DLG es simétrica.

Si $\alpha_3 < 0$, entonces la distribución DLG es asimétrica negativa (hacia la izquierda).

Si $\alpha_3 > 0$, entonces la distribución DLG es asimétrica positiva (hacia la derecha).

Si $\alpha_4 = 3$, entonces la distribución DLG es mesocúrtica.

Si $\alpha_4 < 3$, entonces la distribución DLG es platicúrtica.

Si $\alpha_4 > 3$, entonces la distribución DLG es leptocúrtica.

Con el propósito de familiarizar al lector con el manejo de las distribuciones provenientes de la familia Lambda Generalizada, a continuación se proporcionan algunos ejemplos de funciones de densidad provenientes de la DLG que corresponden a distribuciones que presentan una determinada forma, tal como: simétricas, asimétricas, leptocúrticas, platicúrticas, mesocúrticas o combinaciones de estas. También se indican algunas distribuciones usuales, como la normal estándar y la uniforme en el intervalo (0,1), como distribuciones provenientes de la Distribución Lambda Generalizada.

5.1.3 Distribuciones simétricas provenientes de la DLG

Las distribuciones simétricas provenientes de la DLG se caracterizan porque su momento de orden tres es igual a cero, es decir, $\alpha_3 = 0$; dicho momento se convierte en cero

cuando se tiene que $\lambda_3 = \lambda_4$. Enseguida, se presentan dos ejemplos de densidades correspondientes a distribuciones simétricas obtenidas al asignar valores adecuados y específicos a los cuatro parámetros de la DLG. La primera corresponde a una función de densidad simétrica que además es mesocúrtica $\alpha_4 = 3$, de media $\alpha_1 = 2$ y varianza $\alpha_2 = \sigma^2 = 1$. La segunda es simétrica $\alpha_3 = 0$ y además platicúrtica, puesto que su cuarto momento es de $\alpha_4 = 1.8$ con media $\alpha_1 = 1$ y varianza $\alpha_2 = \sigma^2 = \frac{4}{12}$.

5.1.3.1 Distribución simétrica mesocúrtica. Si

$\lambda_1 = 2, \lambda_2 = 0.1975, \lambda_3 = 0.1349, \lambda_4 = 0.1349$, la DLG se expresa de la siguiente forma:

$$F^{-1}(y) = 2 + \frac{y^{0.1349} - (1-y)^{0.1349}}{0.1975} = x, \text{ donde } 0 \leq y \leq 1. \quad (5.2)$$

su función de densidad es

$$f(x) = \frac{0.1975}{0.1349y^{0.1349-1} + 0.1349(1-y)^{0.1349-1}}, \text{ con } x = F^{-1}(y)$$

Como $0 \leq y \leq 1$, en el caso en el que se tenga $y = 0.1$ resulta

$$F^{-1}(0.1) = 2 + \frac{0.1^{0.1349} - (0.9)^{0.1349}}{0.1975} = 0.72 = x$$

$$f(0.72) = \frac{0.1975}{0.1349(0.1)^{0.1349-1} + 0.1349(0.9)^{0.1349-1}} \cong 0.173$$

Para $y = 0.5$ se obtiene:

$$F^{-1}(0.5) = 2 + \frac{0.5^{0.1349} - (0.5)^{0.1349}}{0.1975} = 2 = x$$

$$f(2) = \frac{0.1975}{0.1349(0.5)^{0.1349-1} + 0.1349(0.5)^{0.1349-1}} \cong 0.4$$

Con $y = 0.9$ los resultados son los siguientes:

$$F^{-1}(0.9) = 2 + \frac{0.9^{0.1349} - (0.1)^{0.1349}}{0.1975} = 2.28 = x$$

$$f(0.72) = \frac{0.1975}{0.1349(0.9)^{0.1349-1} + 0.1349(0.1)^{0.1349-1}} \cong 0.173$$

Continuando con este procedimiento, se conforma la Tabla 5.1.

Tabla 5.1. Valores de y , x , $f(x)$ para la distribución simétrica-mesocúrtica dada en (5.2)

y	0.05	0.1	0.3	0.5	0.7	0.9	0.95
x	0.36	0.72	1.48	2.0	2.52	3.28	3.64
$f(x)$	0.101	0.173	0.348	0.4	0.348	0.173	0.101

Con los valores que se presentan en la Tabla 5.1 se obtiene la curva que se visualiza en la Figura 5.1.

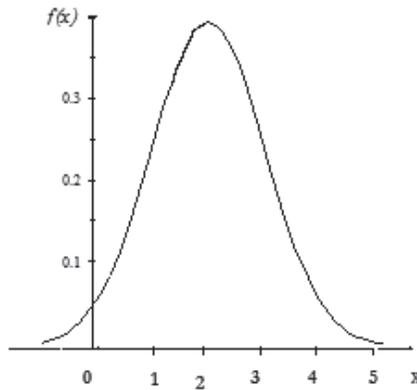


Figura 5.1. Gráfica de la densidad $f(x)$ correspondiente a (5.2)
(Elaborada por los autores con el software libre R)

5.1.3.2 Distribución simétrica platocúrtica. La Distribución Lambda Generalizada con los parámetros que se especifican a continuación, $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 2$, $\lambda_4 = 2$, se puede escribir de la siguiente forma:

$$F^{-1}(y) = 1 + \frac{y^2 - (1-y)^2}{1} = 2y \quad (5.3)$$

donde $0 \leq y \leq 1$ y cero en otro caso; corresponde a la distribución uniforme en el intervalo $[0,2]$.

En el caso de que se asigne

$$y = 0.1; F^{-1}(0.1) = 1 + \frac{0.1^2 - (1-0.1)^2}{1} = 2(0.1) = 0.2 = x$$

la función de densidad

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}}$$

se expresará de la siguiente manera:

$$f(x) = \frac{1}{2y^{2-1} + 2(1-y)^{2-1}} = \frac{1}{2} = 0.5$$

para todo $x \in [0,2]$. En los demás casos la función toma el valor de cero.

$$f(0.2) = \frac{1}{2y^{2-1} + 2(1-y)^{2-1}} = \frac{1}{2} = 0.5$$

$$y = 0.5 \rightarrow F^{-1}(0.5) = 1 + \frac{0.5^2 - (1-0.5)^2}{1} = 2(0.5) = 1.0 = x$$

$$f(1.0) = \frac{1}{2} = 0.5$$

$$y = 0.9 \rightarrow F^{-1}(0.9) = 1 + \frac{0.9^2 - (1-0.9)^2}{1} = 2(0.9) = 1.8 = x$$

$$f(1.8) = \frac{1}{2} = 0.5$$

Siguiendo con el proceso antes indicado, se conforma la Tabla 5.2.

**Tabla 5.2. Valores de y , x , $f(x)$
para la distribución simétrica-platicúrtica dada en (5.3)**

y	0.05	0.1	0.3	0.5	0.7	0.9	0.95
x	0.1	0.2	0.6	1.0	1.4	1.8	1.9
$f(x)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Al representar gráficamente los pares de valores de la forma $(x, f(x))$ en el plano cartesiano, y teniendo en cuenta que $x \in [0,2]$, se obtiene la Figura 5.2, de una función de densidad simétrica $\alpha_3 = 0$ correspondiente a una distribución uniforme con coeficiente de curtosis $\alpha_4 = 1.8$, con media $\alpha_1 = 2$ definida en el intervalo $[0,2]$,

y con varianza $\alpha_2 = \sigma^2 = \frac{(2-0)^2}{12} = \frac{4}{12}$

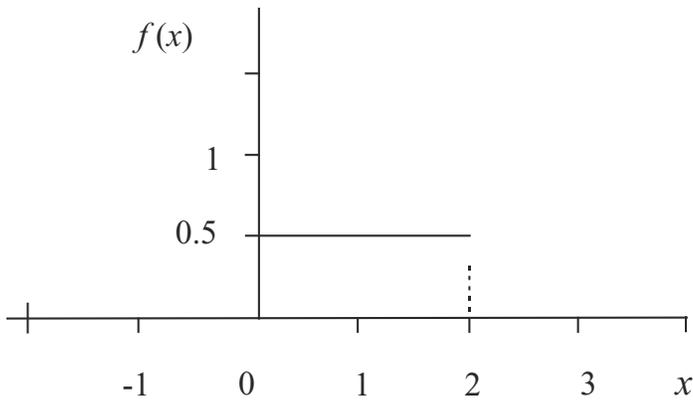


Figura 5.2. Gráfica de la función de densidad $f(x)$ de (5.3)
(Elaborada por los autores con el software libre R).

5.1.4 Distribuciones asimétricas seleccionadas de la DLG

A continuación se presentan dos ejemplos de densidades correspondientes a distribuciones asimétricas obtenidas al asignar valores específicos a los cuatro parámetros de la DLG. La primera corresponde a una función de densidad asimétrica para la derecha con coeficiente de asimetría $\alpha_3 = 0.2$ y la segunda con coeficiente de asimetría $\alpha_3 = 0.45$, también asimétrica para la derecha, porque su coeficiente de asimetría es mayor que cero. Además, para la primera se determinan sus cuatro momentos a manera de ejemplo, y para la segunda se puede proceder de modo similar.

5.1.4.1 Distribución asimétrica platicúrtica. Al considerar los valores siguientes para los cuatro parámetros de la DLG, $\lambda_1 = 2.9896$, $\lambda_2 = 0.2947$, $\lambda_3 = 0.06384$, $\lambda_4 = 0.5571$.

$$F^{-1}(y) = 2.9896 + \frac{y^{0.06384} - (1-y)^{0.5571}}{0.2947} = x \quad \text{donde } 0 \leq y \leq 1 \quad (5.4)$$

La función de densidad de probabilidad resultante es:

$$f(x) = \frac{0.2947}{0.06384y^{0.06384-1} + 0.5571(1-y)^{0.5571-1}} \quad \text{con } x = F^{-1}(y)$$

Como $0 \leq y \leq 1$, al considerar el valor de $y = 0.1$ resultan:

$$F^{-1}(0.1) = 2.9896 + \frac{0.1^{0.06384} - (0.9)^{0.5571}}{0.2947} \cong 2.72 = x$$

$$f(2.72) = \frac{0.2947}{0.06384(0.1)^{0.06384-1} + 0.5571(1-0.1)^{0.5571-1}} \cong 0.259$$

Al realizar los cálculos para los valores de “ y ” que aparecen en la Tabla 5.3 se obtienen los valores para x y $f(x)$.

Tabla 5.3. Valores distribución asimétrica-platicúrtica (5.4)

y	0.001	0.1	0.2	0.5	0.7	0.9	0.999
x	1.789	2.72	3.055	3.929	4.571	5.419	6.309
$f(x)$	0.007	0.259	0.326	0.347	0.283	0.182	0.024

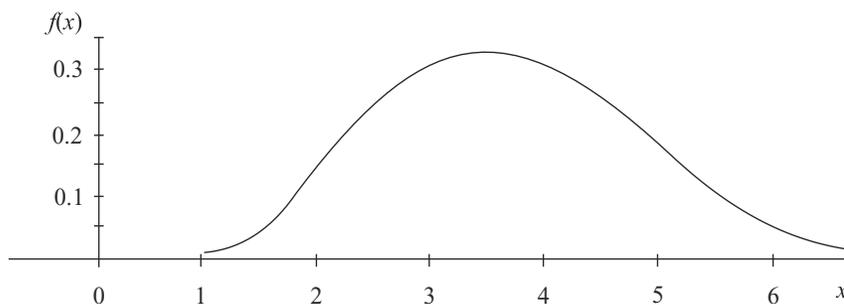


Figura 5.3. Gráfica de la función de densidad $f(x)$ de (5.4)

(Elaborada por los autores con el software libre R)

La curva que se presenta en la Figura 5.3 corresponde a una función de densidad de probabilidad asimétrica para la derecha, puesto que $\alpha_3 = 0.2$ es mayor que cero, y es una curva platicúrtica, ya que $\alpha_4 = 2.2$ es menor que 3; dicha función de densidad presenta una media $\alpha_1 = 4$ y varianza $\alpha_2 = 1$. En efecto, utilizando las expresiones que involucran los cuatro momentos para la distribución Lambda Generalizada y realizando los cálculos a que hay lugar se tiene:

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4} = \frac{1}{1 + 0.06384} - \frac{1}{1 + 0.5571} \cong 0.29777$$

La media se obtiene de la siguiente manera:

$$\alpha_1 = \lambda_1 + \frac{A}{\lambda_2} = 2.9896 + \frac{0.29777}{0.2947} \cong 4$$

Ahora,

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\text{Beta}(1 + \lambda_3, 1 + \lambda_4)$$

Con

$$\text{Beta}(1 + \lambda_3, 1 + \lambda_4) = \text{Beta}(1.06384, 1.5571) = \int_0^1 x^{0.06384} (1-x)^{0.5571} dx = 0.592124$$

$$B = \frac{1}{1 + 2(0.06384)} + \frac{1}{1 + 2(0.5571)} - 2(0.592124) \cong 0.17552$$

La varianza se obtiene realizando los siguientes reemplazos:

$$\alpha_2 = \sigma^2 = \frac{B - A^2}{\lambda_2^2} = \frac{0.17552 - (0.29777)^2}{(0.2947)^2} = 1$$

De la anterior expresión se obtiene que

$$\sigma = 1$$

Por otra parte,

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\text{Beta}(1 + 2\lambda_3, 1 + \lambda_4) + 3\text{Beta}(1 + \lambda_3, 1 + 2\lambda_4)$$

Calculando el valor de la función Beta se obtiene:

$$\text{Beta}(1 + 2\lambda_3, 1 + \lambda_4) = \text{Beta}(1.12768, 1.5571) = \int_0^1 x^{0.12768} (1-x)^{0.5571} dx = 0.54826$$

$$\text{Beta}(1 + \lambda_3, 1 + 2\lambda_4) = \text{Beta}(1.06384, 2.1142) = \int_0^1 x^{0.06384} (1-x)^{1.1142} dx = 0.429658$$

$$C = \frac{1}{1 + 3(0.06384)} - \frac{1}{1 + 3(0.5571)} - 3(0.54826) + 3(0.429658) \cong 0.1091$$

El coeficiente de asimetría se obtiene al realizar la siguiente operatoria:

$$\alpha_3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3} = \frac{0.1091 - 3(0.29777)(0.17552) + 2(0.29777)^3}{(0.2947)^3 (1)^3}$$

$$\alpha_3 = 0.19969 \cong 0.2$$

Además se sabe que

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\text{Beta}(1 + 3\lambda_3, 1 + \lambda_4) + 6\text{Beta}(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\text{Beta}(1 + \lambda_3, 1 + 3\lambda_4)$$

Como

$$\text{Beta}(1 + 3\lambda_3, 1 + \lambda_4) = \int_0^1 x^{0.19152} (1-x)^{0.5571} dx = 0.509583$$

$$\text{Beta}(1 + 2\lambda_3, 1 + 2\lambda_4) = \int_0^1 x^{0.12768} (1-x)^{1.1142} dx = 0.3921$$

$$\text{Beta}(1 + \lambda_3, 1 + 3\lambda_4) = \int_0^1 x^{0.06384} (1-x)^{1.16713} dx = 0.335969$$

se deduce que

$$D = \frac{1}{1 + 4(0.6384)} + \frac{1}{1 + 4(0.5571)} - 4(0.509583) + 6(0.3921) - 4(0.335969)$$

$$D \cong 0.0767272$$

Al realizar las operaciones siguientes, el coeficiente de curtosis resultante es:

$$\alpha_4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4} = \frac{0.0767272 - 4(0.29777)(0.1091) + 6(0.29777)^2(0.17552) - 3(0.29777)^4}{(0.2947)^4 (1)^4}$$

$$\alpha_4 = 2.1971 \cong 2.2$$

5.1.4.2. Distribución asimétrica leptocúrtica. Si se consideran los parámetros siguientes para la DLG: $\lambda_1 = 0.8163$ $\lambda_2 = 0.05419$ $\lambda_3 = 0.02412$ $\lambda_4 = 0.0388$

se obtiene:

$$F^{-1}(y) = 0.8163 + \frac{y^{0.02412} - (1-y)^{0.0388}}{0.05419} = x \quad \text{con } 0 \leq y \leq 1. \quad (5.5)$$

La función de densidad es:

$$f(x) = \frac{0.05419}{0.02412y^{0.02412-1} + 0.0388(1-y)^{0.0388-1}} \quad \text{para } x = F^{-1}(y)$$

Si se especifica el valor de $y = 0.2$ se obtiene:

$$F^{-1}(0.2) = 0.8163 + \frac{0.2^{0.02412} - (1-0.2)^{0.0388}}{0.05419} = 0.273 = x$$

Ahora, el valor de la función de densidad de probabilidad correspondiente es:

$$f(0.273) = \frac{0.05419}{0.02412(0.2)^{0.02412-1} + 0.0388(1-0.2)^{0.388-1}} = 0.33$$

Si se procede de similar forma y se realizan los cálculos pertinentes, es posible conformar la Tabla 5.4.

Tabla 5.4. Valores distribución asimétrica-leptocúrtica (5.5)

y	0.01	0.1	0.2	0.5	0.7	0.9	0.999
x	-1.11	-0.1	0.273	1.00	1.5	2.346	3.83
$f(x)$	0.024	0.2	0.33	0.44	0.343	0.142	0.016

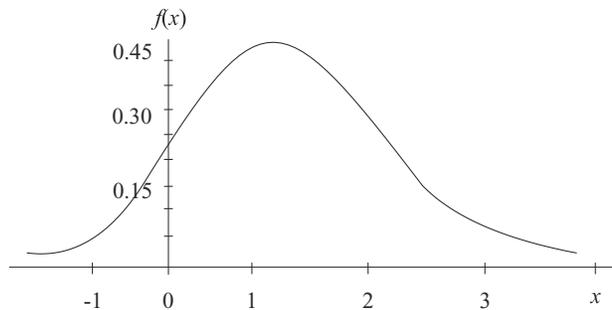


Figura 5.4. Gráfica de la función de densidad $f(x)$ de (5.5)
(Elaborada por los autores con el software libre R)

La curva que se presenta en la Figura 5.4 corresponde a una función de densidad de probabilidad asimétrica para la derecha, puesto que $\alpha_3 = 0.45$ es mayor que cero, y es leptocúrtica, porque su coeficiente de curtosis es $\alpha_4 = 4$, el cual es mayor que tres; la mencionada densidad representa una variable aleatoria con media $\alpha_1 = 1.07$, y varianza $\alpha_2 = 1$, al proceder de manera similar a lo expuesto al respecto en la subsección 5.1.4.1.

5.1.5 Distribuciones usuales provenientes de la DLG

En esta subsección se presentan dos funciones de densidad de probabilidad usuales provenientes de la DLG; se trata de la distribución normal estándar y de la distribución uniforme en el intervalo $(0,1)$; la primera, de gran importancia para el estudio de algunos fenómenos aleatorios y como función límite en procesos de convergencia asintótica, y la segunda es útil para generar números aleatorios con el propósito de simular valores de variables aleatorias con una especificada función de distribución.

5.1.5.1 Distribución normal estándar proveniente de la DLG. La distribución normal estándar se puede aproximar por medio de una distribución Lambda Generalizada si se asumen los siguientes cuatro parámetros: $\lambda_1 = 0$, $\lambda_2 = 0.1975$, $\lambda_3 = 0.1349$, $\lambda_4 = 0.1349$ (Karian & Dudewics, 2000). En este caso, la inversa de la función de distribución toma la siguiente forma:

$$F^{-1}(y) = \frac{y^{0.1349} - (1-y)^{0.1349}}{0.1975} = x, \text{ donde } 0 \leq y \leq 1. \quad (5.6)$$

La función de densidad de probabilidad correspondiente es:

$$f(x) = \frac{0.1975}{0.1349y^{0.1349-1} + 0.1349(1-y)^{0.1349-1}}, \text{ con } x = F^{-1}(y)$$

Puesto que $0 \leq y \leq 1$, si se asume el valor de $y = 0.3$ se obtiene:

$$F^{-1}(0.3) = \frac{0.3^{0.1349} - (0.7)^{0.1349}}{0.1975} = -0.52 = x$$

$$f(-0.52) = \frac{0.1975}{0.1349(0.3)^{0.1349-1} + 0.1349(0.7)^{0.1349-1}} = 0.348$$

Al considerar el valor de $y = 0.5$ resulta:

$$F^{-1}(0.3) = \frac{0.5^{0.1349} - (0.5)^{0.1349}}{0.1975} = 0 = x$$

$$f(0) = \frac{0.1975}{0.1349(0.5)^{0.1349-1} + 0.1349(0.5)^{0.1349-1}} \cong 0.398$$

Si se considera el valor de $y = 0.7$ se obtiene:

$$F^{-1}(0.7) = \frac{0.7^{0.1349} - (0.3)^{0.1349}}{0.1975} = 0.52 = x$$

$$f(0.52) = \frac{0.1975}{0.1349(0.7)^{0.1349-1} + 0.1349(0.3)^{0.1349-1}} = 0.348$$

Siguiendo con el proceso se obtienen los valores de la Tabla 5.5.

Tabla 5.5. Valores distribución normal estándar dada en (5.6)

y	0.05	0.1	0.3	0.5	0.7	0.9	0.95
x	-1.64	-1.28	-0.52	0.0	0.52	1.28	1.64
$f(x)$	0.101	0.17	0.348	0.398	0.348	0.17	0.101

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

La representación de los pares de puntos $(x, f(x))$ en el plano cartesiano (Figura 5.5) es una curva que coincide con la gráfica de la función de densidad de una distribución *normal estándar* $N(0,1)$. El coeficiente de asimetría es $\alpha_3 = 0$, y el coeficiente de curtosis es $\alpha_4 = 3$.

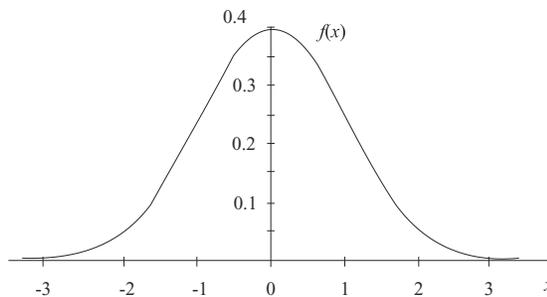


Figura 5.5. Gráfica de la función de densidad $f(x)$ de (5.6)
(Elaborada por los autores con el software libre R)

5.1.5.2 Distribución uniforme en (0,1) proveniente de la DLG. La distribución uniforme en el intervalo (0,1) se puede aproximar por medio de una DLG si se asumen los siguientes cuatro parámetros: $\lambda_1 = 0.5$, $\lambda_2 = 2$, $\lambda_3 = 1$, $\lambda_4 = 1$ (Karian & Dudewics, 2000). La inversa de la función de distribución se puede expresar de la siguiente forma:

$$F^{-1}(y) = 0.5 + \frac{y^2 - (1-y)^2}{2} = y \quad (5.7)$$

donde $0 \leq y \leq 1$ y cero en otro caso. Al asumir que

$$y = 0.1 \rightarrow F^{-1}(0.1) = 0.5 + \frac{0.1^2 - (1-0.1)^2}{2} = 0.1 = x$$

La función de densidad siguiente:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}}$$

se puede expresar así:

$$f(x) = \frac{2}{y^{1-1} + (1-y)^{1-1}} = 1$$

La función anterior es igual a 1 para todo $x \in (0,1)$; en los demás casos la función toma el valor de cero. Luego

$$f(0.1) = \frac{2}{(0.1)^{1-1} + (1-0.1)^{1-1}} = 1$$

Si

$$y = 0.5 \rightarrow F^{-1}(0.5) = 0.5 + \frac{0.5^2 - (1-0.5)^2}{2} = 0.5 = x$$

Con lo cual

$$f(0.5) = 1$$

Ahora, si

$$y = 0.9 \rightarrow F^{-1}(0.9) = 0.5 + \frac{0.9^2 - (1-0.9)^2}{2} = 0.9 = x$$

Con esto,

$$f(0.9) = 1$$

Al realizar un proceso similar al anterior se logran los valores que se presentan en la Tabla 5.6.

Tabla 5.6. Valores distribución uniforme en (0,1) dada en (5.7)

y	0.05	0.1	0.3	0.5	0.7	0.9	0.95
x	0.05	0.1	0.3	0.5	0.7	0.9	0.95
$f(x)$	1	1	1	1	1	1	1

Al graficar los pares de valores $(x, f(x))$, teniendo en cuenta que $x \in (0,1)$, se obtiene la Figura 5.6 de una función de densidad simétrica $\alpha_3 = 0$, que corresponde a una distribución uniforme definida en el intervalo $(0,1)$, con coeficiente de curtosis $\alpha_4 = 1.8$, con media $\alpha_1 = 0.5$ y con varianza

$$\alpha_2 = \sigma^2 = \frac{(1-0)^2}{12} = \frac{1}{12}$$

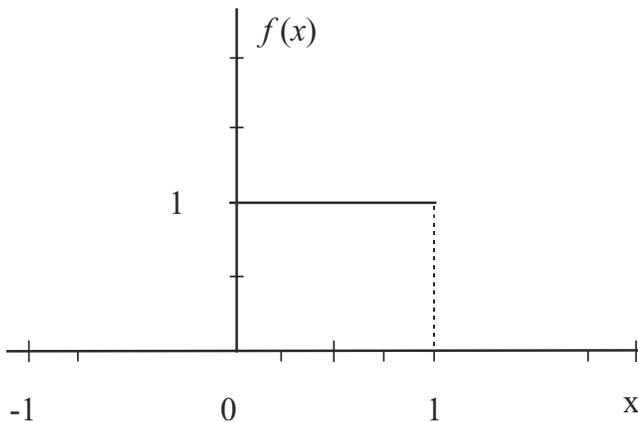


Figura 5.6. Gráfica de la función de densidad $f(x)$ de (5.7)
(Elaborada por los autores con el software libre R)

5.1.6 Otras distribuciones provenientes de la DLG

Con los parámetros $\lambda_1 = 0.9935$, $\lambda_2 = 1.0491$, $\lambda_3 = 0.2121$, $\lambda_4 = 0.1061$ para la Distribución Lambda Generalizada, se obtiene un momento de orden cuatro dado por $\alpha_4 = 2.88$, el cual caracteriza a una distribución platicúrtica proveniente de la DLG; dicha distribución se puede expresar de la siguiente manera:

$$F^{-1}(y) = 0.9935 + \frac{y^{0.2121} - (1-y)^{0.1061}}{1.0491} = x \quad (5.8)$$

La función de densidad de probabilidad correspondiente es:

$$f(x) = \frac{1.0491}{0.2121y^{0.2121-1} + 0.1061(1-y)^{0.1061-1}}$$

Puesto que $\alpha_3 = 0.2541$ es menor que cero, se tiene una distribución asimétrica para la izquierda con media $\alpha_1 = 0.9182$ y con varianza $\alpha_2 = \sigma^2 = 0.0442$. La anterior función de densidad corresponde a una distribución de Weibull de parámetros 1 y 5, respectivamente (Karian & Dudewics, 2000), útil para modelar, por ejemplo, funciones de supervivencia (Marshall, 2007), tasas de fallos y riesgos.

Por otro lado, si $\lambda_1 = -0.03767$, $\lambda_2 = -0.2807$, $\lambda_3 = -0.1184$, $\lambda_4 = -0.1265$, entonces se obtiene el valor $\alpha_4 = 8.1$ para el momento de orden cuatro, el cual es mayor que tres, indicando que se trata de una distribución leptocúrtica proveniente de la DLG, cuya inversa de la función de distribución es:

$$F^{-1}(y) = -0.03767 + \frac{y^{-0.1184} - (1-y)^{-0.1265}}{-0.2807} = x \quad (5.9)$$

La función de densidad correspondiente es:

$$f(x) = \frac{-0.2807}{-0.1184y^{-0.1184-1} - 0.1265(1-y)^{-0.1265-1}}$$

Finalmente, si $\lambda_1 = 0.8451$, $\lambda_2 = 0.1085$, $\lambda_3 = 0.01017$, $\lambda_4 = 0.03422$, entonces, se obtiene el valor $\alpha_4 = 5.09$ para el momento de orden cuatro, el cual es mayor que tres, indicando que se trata nuevamente de una distribución leptocúrtica proveniente de la Distribución Lambda Generalizada, cuya inversa de la función de distribución es:

$$F^{-1}(y) = 0.8451 + \frac{y^{0.01017} - (1-y)^{0.03422}}{0.1085} = x \quad (5.10)$$

La función de densidad correspondiente es:

$$f(x) = \frac{0.1085}{0.01017y^{0.01017-1} + 0.03422(1-y)^{0.03422-1}}$$

Debido a que $\alpha_3 = 1.0687$ es mayor que cero, resulta una distribución asimétrica para la derecha con media $\alpha_1 = 1.0571$ y con varianza $\alpha_2 = \sigma^2 = 0.1313$. La anterior distribución guarda mucha similitud con una distribución de lognormal con media cero y desviación estándar de un tercio, respectivamente (Karian & Dudewics, 2000), requerida en el modelamiento de tasas de fallas y riesgos.

5.2 Simulación de variables aleatorias con distribución simétrica y asimétrica proveniente de la DLG

En esta sección se presenta un algoritmo general basado en el método de la transformada inversa para generar valores de variables aleatorias con distribuciones provenientes de la DLG, caracterizadas por ser simétricas, asimétricas, platicúrticas o leptocúrticas. Para aplicar dicho algoritmo se requiere generar números aleatorios provenientes de una distribución uniforme en el intervalo (0,1) que serán utilizados en cada caso.

Con el propósito de orientar al lector, se proporcionan ejemplos de simulación en cuatro grupos, así: distribuciones simétricas, asimétricas, platicúrticas y leptocúrticas, incluyendo algunas distribuciones usuales; se indican algunos ejemplos de los cálculos involucrados y se elaboran algunas tablas con los valores simulados.

5.2.1 Algoritmo general para generar valores de una variable con distribución proveniente de la DLG

En términos generales, para obtener valores de una variable aleatoria continua proveniente de la DLG también es posible utilizar el método de la transformada inversa descrito en la sección 2.7.1. En esencia, se usa la función percentil de la DLG, dada en la expresión (5.1). Los valores simulados para la variable aleatoria X con DLG se obtienen generando números aleatorios u con distribución $U(0,1)$, los cuales son reemplazados adecuadamente en la expresión (5.1) para obtener como resultado la expresión (5.11):

$$x = F^{-1}(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \quad (5.11)$$

El algoritmo general consta de los siguientes pasos:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Reemplazar en

$$x = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2}$$

3. Regresar al paso 1 y repetir el procedimiento hasta completar la cantidad de valores x que se requieran para la variable aleatoria X .

En los ejemplos que se desarrollan en las siguientes subsecciones se utilizan frecuentemente los números aleatorios de la Tabla 4.5.

5.2.2 Simulación de valores de variables aleatorias con distribución simétrica

Utilizando adecuadamente la expresión (5.11) y el algoritmo descrito anteriormente, se obtienen los valores simulados x de la variable X con distribución simétrica mesocúrtica dada en la expresión (5.2) mediante la expresión (5.12) que se indica a continuación:

$$x = F^{-1}(u) = 2 + \frac{u^{0.1349} - (1-u)^{0.1349}}{0.1975} \quad (5.12)$$

El algoritmo resultante es el siguiente:

1. Generar un número aleatorio x con distribución uniforme $U(0,1)$
2. Reemplazar en

$$x = 2 + \frac{u^{0.1349} - (1-u)^{0.1349}}{0.1975}$$

A fin de describir la forma como se obtienen los valores simulados, a continuación se determinan los primeros 3 valores x de la variable aleatoria X correspondiente a la expresión (5.12), utilizando los 3 primeros números aleatorios de la Tabla 4.5.

$$u = 0.559355 \rightarrow$$

$$x = 2 + \frac{0.559355^{0.1349} - (1 - 0.559355)^{0.1349}}{0.1975} = 2.148256$$

$$u = 0.982616 \rightarrow$$

$$x = 2 + \frac{0.982616^{0.1349} - (1 - 0.982616)^{0.1349}}{0.1975} = 4.120229$$

$$u = 0.325262 \rightarrow$$

$$x = 2 + \frac{0.325262^{0.1349} - (1 - 0.325262)^{0.1349}}{0.1975} = 1.549871$$

Se procede de manera similar para los restantes 45 números aleatorios. Los valores que se presentan en la Tabla 5.7 se obtuvieron utilizando un programa de computador para automatizar los cálculos.

Realizando un proceso similar, los valores simulados x de la variable X con distribución simétrica platocúrtica definida por la expresión (5.3) se obtienen usando la expresión (5.13):

$$x = F^{-1}(u) = 1 + \frac{u^2 - (1-u)^2}{1} = 2u \quad (5.13)$$

Tabla 5.7. Valores simulados x de la variable aleatoria X , expresión 5.12

N.º	u	x	N.º	u	x
1	0,559355	2,148256	26	0,761206	2,706543
2	0,982616	4,120229	27	0,803134	2,84928
3	0,325262	1,549871	28	0,702858	2,529403
4	0,535503	2,08846	29	0,047913	0,330803
5	0,507678	2,019105	30	0,394261	1,733648
6	0,281995	1,426457	31	0,22862	1,260265
7	0,347206	1,609733	32	0,726215	2,597957
8	0,921085	3,412776	33	0,882465	3,185467
9	0,408148	1,769335	34	0,182222	1,096542
10	0,095739	0,694602	35	0,27969	1,419638
11	0,333891	1,573599	36	0,695545	2,508475
12	0,620771	2,305383	37	0,902572	3,295419
13	0,641131	2,359046	38	0,122813	0,841107
14	0,752135	2,677615	39	0,507697	2,019153
15	0,204952	1,179569	40	0,117283	0,813245
16	0,143259	0,937001	41	0,924725	3,438269
17	0,445881	1,86491	42	0,517283	2,043018
18	0,369064	1,667949	43	0,271406	1,394901
19	0,48528	1,963364	44	0,644636	2,368389
20	0,099268	0,715297	45	0,643786	2,366119
21	0,725651	2,596266	46	0,852829	3,045753
22	0,037251	0,211051	47	0,06515	0,485436
23	0,174697	1,067621	48	0,406798	1,76588
24	0,288352	1,445116	49	0,566701	2,166772

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

En este caso, el algoritmo que se ha de seguir es:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Sustituir en

$$x = 1 + \frac{u^2 - (1-u)^2}{1} = 2u$$

Tabla 5.8. Valores simulados x de la variable aleatoria X , expresión (5.13)

N.º	u	x	N.º	u	x
1	0,559355	1,118711	26	0,761206	1,522411
2	0,982616	1,965231	27	0,803134	1,606269
3	0,325262	0,650523	28	0,702858	1,405716
4	0,535503	1,071005	29	0,047913	0,095826
5	0,507678	1,015355	30	0,394261	0,788522
6	0,281995	0,563991	31	0,22862	0,457241
7	0,347206	0,694412	32	0,726215	1,452431
8	0,921085	1,842169	33	0,882465	1,76493
9	0,408148	0,816296	34	0,182222	0,364444
10	0,095739	0,191477	35	0,27969	0,559379
11	0,333891	0,667782	36	0,695545	1,391091
12	0,620771	1,241542	37	0,902572	1,805143
13	0,641131	1,282262	38	0,122813	0,245627
14	0,752135	1,504271	39	0,507697	1,015394
15	0,204952	0,409904	40	0,117283	0,234566
16	0,143259	0,286519	41	0,924725	1,84945
17	0,445881	0,891761	42	0,517283	1,034566
18	0,369064	0,738127	43	0,271406	0,542812
19	0,48528	0,97056	44	0,644636	1,289273
20	0,099268	0,198535	45	0,643786	1,287571
21	0,725651	1,451302	46	0,852829	1,705657
22	0,037251	0,074503	47	0,06515	0,1303
23	0,174697	0,349395	48	0,406798	0,813596
24	0,288352	0,576703	49	0,566701	1,133402
25	0,66887	1,337741	50	0,84329	1,68658

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

En la Tabla 5.8 se presentan 50 valores simulados mediante la expresión (5.13), utilizando los números aleatorios de la Tabla 4.5.

Por otro lado, para generar los valores simulados de la variable con distribución normal estándar definida en (5.6), la cual también es simétrica con respecto al origen, se usa la expresión (5.14).

$$z = F^{-1}(u) = \frac{u^{0.1349} - (1-u)^{0.1349}}{0.1975} \quad (5.14)$$

Para obtener los valores simulados se procede como se ejemplificó en la sección 4.1, subsección 4.1.4.

Finalmente, los valores simulados x de la variable X con distribución uniforme en el intervalo $(0,1)$, definida en (5.7), la cual es simétrica con respecto a su media igual a 0.5, se obtienen de la expresión (5.15).

$$x = F^{-1}(u) = 0.5 + \frac{u^2 - (1-u)^2}{2} = u \quad (5.15)$$

Ahora, el algoritmo que se ha de seguir es:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Sustituir en

$$x = 0.5 + \frac{u^2 - (1-u)^2}{2} = u$$

En este caso, los valores x simulados coinciden con los valores u de una variable aleatoria con distribución uniforme en el intervalo $(0,1)$, por ejemplo, los u correspondientes a la Tabla 4.5.

5.2.3 Simulación de variables aleatorias con distribución asimétrica

Para generar valores de una variable aleatoria con distribución asimétrica proveniente de la DLG se utilizan la expresión (5.11) y su algoritmo general correspondiente. Para obtener los valores simulados x de la variable X con distribución asimétrica platicúrtica definida en la expresión (5.4) se usa la expresión (5.16).

$$x = F^{-1}(u) = 2.9896 + \frac{u^{0.06384} - (1-u)^{0.5571}}{0.2947} \quad (5.16)$$

El algoritmo resultante es el siguiente:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Reemplazar en

$$x = 2.9896 + \frac{u^{0.06384} - (1-u)^{0.5571}}{0.2947}$$

Para ilustrar el procedimiento, se calculan los dos primeros valores simulados por medio de los dos primeros números aleatorios indicados en la Tabla 4.5.

$$u = 0.559355 \rightarrow$$

$$x = 2.9896 + \frac{0.559355^{0.06384} - (1 - 0.559355)^{0.5571}}{0.2947} = 4.109812$$

$$u = 0.982616 \rightarrow$$

$$x = 2.9896 + \frac{0.982616^{0.06384} - (1 - 0.982616)^{0.5571}}{0.2947} = 6.024098$$

Como resultado del proceso de simulación utilizando el algoritmo indicado anteriormente, enseguida se usan los 50 números aleatorios presentados en la Tabla 4.5 para generar 50 valores x de las variables aleatorias X por medio de la expresión (5.16). Los valores se presentan en la Tabla 5.9.

Tabla 5.9. Valores simulados x de la variable aleatoria X asimétrica platicúrtica usando (5.16)

N.º	u	x	N.º	u	x
1	0,559355	4,109812	26	0,761206	4,796308
2	0,982616	6,024098	27	0,803134	4,963569
3	0,325262	3,422693	28	0,702858	4,581481
4	0,535503	4,036664	29	0,047913	2,482864
5	0,507678	3,952656	30	0,394261	3,62069
6	0,281995	3,298024	31	0,22862	3,141391
7	0,347206	3,485627	32	0,726215	4,66536
8	0,921085	5,540549	33	0,882465	5,326442
9	0,408148	3,660711	34	0,182222	2,999868
10	0,095739	2,702603	35	0,27969	3,291341
11	0,333891	3,447452	36	0,695545	4,555732
12	0,620771	4,304064	37	0,902572	5,433456
13	0,641131	4,370708	38	0,122813	2,803294
14	0,752135	4,761692	39	0,507697	3,952713
15	0,204952	3,070063	40	0,117283	2,783513
16	0,143259	2,873791	41	0,924725	5,562811
17	0,445881	3,770145	42	0,517283	3,981507
18	0,369064	3,5483	43	0,271406	3,267279
19	0,48528	3,885929	44	0,644636	4,382311
20	0,099268	2,716342	45	0,643786	4,379491
21	0,725651	4,663302	46	0,852829	5,181721
22	0,037251	2,417766	47	0,06515	2,571677
23	0,174697	2,976166	48	0,406798	3,656816
24	0,288352	3,316423	49	0,566701	4,132574
25	0,66887	4,463667	50	0,84329	5,137766

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

Ahora, los valores simulados x de la variable X con distribución asimétrica leptocúrtica definida en (5.5) se obtienen usando la expresión (5.17).

$$x = F^{-1}(u) = 0.8163 + \frac{u^{0.02412} - (1-u)^{0.0388}}{0.05419} \quad (5.17)$$

El algoritmo que se requiere es:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Sustituir en

$$x = 0.8163 + \frac{u^{0.02412} - (1-u)^{0.0388}}{0.05419}$$

Los valores obtenidos a partir de la expresión (5.17) se muestran en la Tabla 5.10.

Tabla 5.10. Valores simulados x de la variable aleatoria X asimétrica leptocúrtica con la expresión (5.17)

N.º	u	x	N.º	u	x
1	0,559355	1,137056	26	0,761206	1,692703
2	0,982616	3,493273	27	0,803134	1,846712
3	0,325262	0,602663	28	0,702858	1,508781
4	0,535503	1,081337	29	0,047913	-0,452594
5	0,507678	1,017477	30	0,394261	0,762108
6	0,281995	0,497055	31	0,22862	0,355943
7	0,347206	0,654268	32	0,726215	1,579048
8	0,921085	2,511241	33	0,882465	2,231753
9	0,408148	0,793459	34	0,182222	0,217324
10	0,095739	-0,127053	35	0,27969	0,491245
11	0,333891	0,623085	36	0,695545	1,487551
12	0,620771	1,286642	37	0,902572	2,36493
13	0,641131	1,338859	38	0,122813	-0,000318
14	0,752135	1,662135	39	0,507697	1,017521
15	0,204952	0,287628	40	0,117283	-0,024281
16	0,143259	0,081756	41	0,924725	2,543607
17	0,445881	0,878174	42	0,517283	1,039406
18	0,369064	0,704746	43	0,271406	0,470187
19	0,48528	0,966713	44	0,644636	1,348012

(Pasa Tabla 5.10)

(Viene Tabla 5.10)

N.º	u	x	N.º	u	x
20	0,09268	-0,109033	45	0,643786	1,345786
21	0,725651	1,577302	46	0,852829	2,067784
22	0,037251	-0,564351	47	0,06515	-0,31195
23	0,174697	0,19281	48	0,406798	0,790418
24	0,288352	0,512965	49	0,566701	1,154439
25	0,66887	1,412789	50	0,84329	2,021006

(Los valores fueron generados por los autores con el software libre R)

5.2.4 Simulación para variable aleatoria con distribución platicúrtica

Para generar valores de una variable aleatoria con distribución platicúrtica proveniente de la DLG se utiliza la expresión (5.11) y su respectivo algoritmo general. Para obtener los valores simulados x de la variable X con distribución platicúrtica definida en la expresión (5.8) se utiliza la expresión (5.18).

$$x = F^{-1}(u) = 0.9935 + \frac{u^{0.2121} - (1-u)^{0.1061}}{1.0491} \quad (5.18)$$

El algoritmo resultante es el siguiente:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Reemplazar en

$$x = 0.9935 + \frac{u^{0.2121} - (1-u)^{0.1061}}{1.0491}$$

Usando los 50 números aleatorios presentados en la Tabla 4.5 se generan los 50 valores x de las variables aleatorias X con distribución de Weibull correspondientes a la expresión (5.18). Los valores se presentan en la Tabla 5.11.

Tabla 5.11. Valores simulados x de la variable aleatoria X de Weibull, platicúrtica usando (5.18)

N.º	u	x	N.º	u	x
1	0,559355	0,962372	26	0,761206	1,074275
2	0,982616	1,323056	27	0,803134	1,101168
3	0,325262	0,830424	28	0,702858	1,039972
4	0,535503	0,949717	29	0,047913	0,545643
5	0,507678	0,934881	30	0,394261	0,872112
6	0,281995	0,801975	31	0,22862	0,763226
7	0,347206	0,8441	32	0,726215	1,053375
8	0,921085	1,202155	33	0,882465	1,162251
9	0,408148	0,8801	34	0,182222	0,724719
10	0,095739	0,62995	35	0,27969	0,800394
11	0,333891	0,835855	36	0,695545	1,035848
12	0,620771	0,995005	37	0,902572	1,181667
13	0,641131	1,005943	38	0,122813	0,664422
14	0,752135	1,068745	39	0,507697	0,934891
15	0,204952	0,744276	40	0,117283	0,657853
16	0,143259	0,687055	41	0,924725	1,206578
17	0,445881	0,901301	42	0,517283	0,940015
18	0,369064	0,857312	43	0,271406	0,794652
19	0,48528	0,922834	44	0,644636	1,007836
20	0,099268	0,634808	45	0,643786	1,007376
21	0,725651	1,053046	46	0,852829	1,137212
22	0,037251	0,518509	47	0,06515	0,581181
23	0,174697	0,717897	48	0,406798	0,879328
24	0,288352	0,806296	49	0,566701	0,966264

(Los valores de esta tabla fueron generados por los autores con el software libre R)

5.2.5 Simulación con variable aleatoria de distribución leptocúrtica

Con el propósito de generar valores de una variable aleatoria con distribución leptocúrtica proveniente de la DLG se utiliza la expresión (5.11) y su respectivo algoritmo general. Para obtener los valores simulados x de la variable X con distribución leptocúrtica dada en la expresión (5.9), se usa la expresión (5.19), como se indica a continuación:

$$x = F^{-1}(u) = -0.03767 + \frac{u^{-0.1184} - (1-u)^{-0.1265}}{-0.2807} \quad (5.19)$$

El algoritmo que se ha de aplicar es el siguiente:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Sustituir en

$$x = -0.03767 + \frac{u^{-0.1184} - (1-u)^{-0.1265}}{-0.2807}$$

Con los 50 números aleatorios presentados en la Tabla 4.5 se generan los 50 valores x de las variables aleatorias X con distribución leptocúrtica concordante con la expresión (5.19). Los valores se presentan en la Tabla 5.12.

Con un proceso similar, los valores simulados x de la variable X con distribución leptocúrtica dada en (5.10) con distribución lognormal se obtienen a través de la expresión (5.20). Los valores se presentan en la Tabla 5.13.

$$x = F^{-1}(u) = 0.8451 + \frac{u^{0.01017} - (1-u)^{0.03422}}{0.1085} \quad (5.20)$$

El algoritmo resultante es el siguiente:

1. Generar un número aleatorio u con distribución uniforme $U(0,1)$
2. Reemplazar en

$$x = 0.8451 + \frac{u^{0.01017} - (1-u)^{0.03422}}{0.1085}$$

Tabla 5.12. Valores simulados x de la variable aleatoria X leptocúrtica con la expresión (5.19)

N.º	u	x	N.º	u	x
1	0,559355	0,097796	26	0,761206	0,552934
2	0,982616	2,340493	27	0,803134	0,681799
3	0,325262	-0,362561	28	0,702858	0,40156
4	0,535503	0,051789	29	0,047913	-1,55791
5	0,507678	-0,001305	30	0,394261	-0,219473
6	0,281995	-0,461238	31	0,22862	-0,598888
7	0,347206	-0,315533	32	0,726215	0,45915
8	0,921085	1,277059	33	0,882465	1,01737
9	0,408148	-0,192051	34	0,182222	-0,741481
10	0,095739	-1,13274	35	0,27969	-0,466769
11	0,333891	-0,343862	36	0,695545	0,384203
12	0,620771	0,220327	37	0,902572	1,13918
13	0,641131	0,262915	38	0,122813	-0,982181
14	0,752135	0,527618	39	0,507697	-0,001269
15	0,204952	-0,66819	40	0,117283	-1,01004
16	0,143259	-0,888854	41	0,924725	1,308173
17	0,445881	-0,118939	42	0,517283	0,016976
18	0,369064	-0,270217	43	0,271406	-0,48691
19	0,48528	-0,043851	44	0,644636	0,270376
20	0,099268	-1,11084	45	0,643786	0,268562
21	0,725651	0,457715	46	0,852829	0,871729
22	0,037251	-1,71738	47	0,06515	-1,36721
23	0,174697	-0,767527	48	0,406798	-0,194702
24	0,288352	-0,446149	49	0,566701	0,112098
25	0,66887	0,323182	50	0,84329	0,830971

(Los valores fueron generados por los autores con el software libre R)

Tabla 5.13. Valores simulados x de la variable aleatoria X lognormal, leptocúrtica, expresión (5.20)

N.º	u	x	N.º	u	x
1	0,559355	1,045683	26	0,761206	1,26036
2	0,982616	2,03683	27	0,803134	1,323165
3	0,325262	0,863678	28	0,702858	1,187013
4	0,535503	1,025442	29	0,047913	0,580132
5	0,507678	1,002582	30	0,394261	0,91503
6	0,281995	0,8311	31	0,22862	0,789317
7	0,347206	0,880013	32	0,726215	1,214802
8	0,921085	1,604486	33	0,882465	1,484504
9	0,408148	0,925432	34	0,182222	0,750119
10	0,095739	0,659479	35	0,27969	0,829341
11	0,333891	0,87011	36	0,695545	1,178677
12	0,620771	1,101308	37	0,902572	1,54143
13	0,641131	1,121139	38	0,122813	0,69185
14	0,752135	1,248037	39	0,507697	1,002597
15	0,204952	0,76978	40	0,117283	0,68565
16	0,143259	0,713393	41	0,924725	1,6185
17	0,445881	0,954033	42	0,517283	1,01039
18	0,369064	0,896256	43	0,271406	0,822992
19	0,48528	0,984676	44	0,644636	1,124636
20	0,099268	0,664021	45	0,643786	1,123785
21	0,725651	1,214108	46	0,852829	1,415143
22	0,037251	0,553781	47	0,06515	0,613856
23	0,174697	0,743365	48	0,406798	0,924419
24	0,288352	0,835936	49	0,566701	1,052052
25	0,66887	1,149557	50	0,84329	1,395521

(Los valores de esta tabla fueron generados por los autores con el software libre R)

5.3 Ejemplos de aplicación de la simulación en contextos reales

A continuación se proporcionan ejemplos prácticos, en los cuales se aplica la simulación en contextos reales; estos permiten observar las potencialidades de los procesos de simulación en diversas disciplinas y en diferentes procesos de investigación. Así mismo, dejan ver que la simulación se constituye en una estrategia para simplificar, complementar y dinamizar los procesos de investigación en cualquier rama del conocimiento humano.

5.3.1 Ejemplo acerca de la variabilidad del espesor de un producto

Una muestra aleatoria de 25 barras de chocolate fueron seleccionadas de un lote en un día particular de la semana, y con base en dicha información se estableció que el modelo apropiado para describir la variabilidad del espesor de las barras de chocolate es un modelo de probabilidad uniforme, según lo ha establecido el departamento de control de calidad de la empresa AA. Se ha estimado que el espesor de las barras de chocolate producidas ese día se encuentra entre 5 y 9 mm. Se requiere simular el espesor de 30 barras de chocolate.

En concordancia con lo expuesto en la subsección 3.2.1, para este caso se considera una variable aleatoria X , que representa el espesor de las barras de chocolate; esta variable sigue una distribución uniforme en el intervalo (a, b) , donde $a = 5$ y $b = 9$, luego los 30 valores simulados x acerca del espesor de las barras de chocolate se pueden obtener mediante la expresión

$$x = a + u(b - a)$$

Es decir,

$$x = 5 + u(9 - 5) = 5 + 4u$$

Utilizando los primeros 30 números aleatorios u de la Tabla 4.6 se obtienen los valores simulados x correspondientes al espesor de las barras de chocolate, los cuales se presentan en la Tabla 5.14.

Tabla 5.14. Valores simulados x de la variable aleatoria X espesor de las barras de chocolate

u	$x = 5+4u$	u	$x = 5+4u$	u	$x = 5+4u$
0,559355	7,23742	0,333891	6,33556	0,725651	7,9026
0,982616	8,93046	0,620771	7,48308	0,037251	5,149
0,325262	6,30105	0,641131	7,56452	0,174697	5,69879
0,535503	7,14201	0,752135	8,00854	0,288352	6,15341
0,507678	7,03071	0,204952	5,81981	0,66887	7,67548
0,281995	6,12798	0,143259	5,57304	0,761206	8,04482
0,347206	6,38882	0,445881	6,78352	0,803134	8,21254
0,921085	8,68434	0,369064	6,47626	0,702858	7,81143
0,408148	6,63259	0,48528	6,94112	0,047913	5,19165
0,095737	5,38295	0,099268	5,39707	0,394261	6,57704

(Los valores de esta tabla fueron generados por los autores con el software libre R)

2.3.2 Tiempos de servicio para los clientes de un supermercado

En el supermercado BB se ha tomado una muestra aleatoria de 100 clientes, para los cuales se ha registrado el tiempo de servicio en la caja registradora n.º 2. El departamento de planeación ha establecido que los tiempos de servicio se ajustan a una distribución exponencial de parámetro $\lambda = 0.1$. Se requiere simular el tiempo de servicio para los siguientes 50 clientes que lleguen a la caja registradora n.º 2.

Utilizando los elementos conceptuales y el algoritmo expuesto en la subsección 3.2.2, aquí se considera una variable aleatoria X para representar los tiempos de servicio en la caja registradora n.º 2. Puesto que esta variable se ajusta a una distribución exponencial de parámetro $\lambda = 0.1$, entonces los 50 valores simulados x referentes a los tiempos de servicio se pueden obtener mediante la expresión

$$x = \frac{-1}{\lambda} \text{Ln}(1-u) = \frac{-1}{0.1} \text{Ln}(1-u)$$

La cual puede expresarse de la siguiente forma:

$$x = -10\text{Ln}(1-u)$$

Tabla 5.15. Valores simulados x de la variable aleatoria X tiempos de servicio en la caja registradora N.º 2

u	x	u	x	u	x
0,559355	8,19516	0,369064	4,60551	0,27969	3,28074
0,982616	40,52205	0,48528	6,64132	0,695545	11,89232
0,325262	3,93431	0,099268	1,04548	0,902572	23,28642
0,535503	7,668	0,725651	12,93354	0,122813	1,31035
0,507678	7,08622	0,037251	0,37963	0,507697	7,08661
0,281995	3,31279	0,174697	1,92005	0,117283	1,24751
0,347206	4,26494	0,288352	3,40172	0,924725	25,86607
0,921085	25,39384	0,66887	11,05244	0,517283	7,28325
0,408148	5,24499	0,761206	14,32154	0,271406	3,16639
0,095737	1,00635	0,803134	16,25232	0,644636	10,34613
0,333891	4,06302	0,702858	12,13545	0,643786	10,32224
0,620771	9,69615	0,047913	0,49099	0,852829	19,1616
0,641131	10,24798	0,394261	5,01306	0,06515	0,67369
0,752135	13,94871	0,22862	2,59574	0,406798	5,2222
0,204952	2,29353	0,726215	12,95412	0,566701	8,36327
0,143259	1,5462	0,882465	21,41019	0,84329	18,53358
0,445881	5,90376	0,182222	2,01164		

(Los valores de esta tabla fueron generados por los autores con el software libre R)

Empleando los 50 números aleatorios u de la Tabla 4.6 se obtienen los valores simulados x correspondientes a los tiempos de servicio en la caja registradora n.º 2., los cuales se pueden observar en la Tabla 5.15. Por medio del valor esperado de la distribución exponencial, se puede establecer que el tiempo promedio de servicio en la caja registradora n.º 2 es de 10 minutos.

$$E(X) = \frac{1}{\lambda} = \frac{1}{0.1} = 10$$

Con los valores simulados se puede establecer que el tiempo promedio de servicio es de 8.81 minutos. De la Tabla 5.15 se observa que los tiempos de servicios podrían estar entre 0.37963 minutos y 40.52205 minutos; es decir, los clientes podrían gastar desde menos de 1 minutos hasta 40 minutos, aproximadamente, en ser atendidos.

5.3.3 Utilidades generadas por unas pequeñas empresas

Las utilidades de las pequeñas empresas de la región T en Colombia, en el mes de mayo del año 2013, se distribuyeron normalmente con media de \$2.500.000 por mes y una

desviación estándar de \$200.000. Se requiere simular las utilidades de 24 pequeñas empresas de la citada región para el mes de mayo de 2014, bajo el supuesto de que se tendrán condiciones similares a las presentadas en mayo de 2013.

En este caso se pueden utilizar los conceptos y resultados indicados en la subsección 4.1.4, y mediante la expresión (4.6) se generarán 24 valores de una variable z normal estándar $N(0,1)$. Para ello se utilizarán los primeros 24 números aleatorios u de la Tabla 4.6.

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

Luego, por medio de la transformación que se indica a continuación, se obtienen los valores simulados x referentes a las utilidades de las pequeñas empresas para el mes de mayo de 2014.

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

Por ejemplo, para $u = 0.559355$ y $z = 0.14825$, con los cuales el valor simulado es:

$$x = z\sigma + \mu = 0.14825(200000) + 2500000 = 2529650.96$$

Los 24 valores x simulados, referentes a las utilidades de las pequeñas empresas para el mes de mayo del año 2014, se presentan en la Tabla 5.16.

Tabla 5.16. Valores simulados x de la variable aleatoria X utilidades de unas pequeñas empresas en la región T de Colombia para mayo de 2014

u	x	u	x	u	x
0,559355	2529651	0,408148	2453867	0,445881	2472982
0,982616	2924047	0,095737	2238919	0,369064	2433590
0,325262	2409974	0,333891	2414720	0,48528	2492673
0,535503	2517692	0,620771	2561077	0,099268	2243060
0,507678	2503821	0,641131	2571809	0,725651	2619253
0,281995	2385291	0,752135	2635523	0,037251	2142209
0,347206	2421947	0,204952	2335914	0,174697	2313524
0,921085	2782556	0,143259	2287400	0,288352	2389023

(Los valores de esta tabla fueron generados por los autores con el software libre R)

Tabla 5.17. X tiempos entre llegadas de 50 clientes en el Banco WW

Tiempo entre llegadas	Frecuencia	Frecuencia relativa	Frecuencia relativa acumulada
2	1	0,02	0,02
4	3	0,06	0,08
6	6	0,12	0,2
8	12	0,24	0,48
10	10	0,2	0,62
12	7	0,14	0,76
14	5	0,1	0,86
16	4	0,08	0,96
18	2	0,04	1
Total	50		

(Los valores de esta tabla fueron generados por los autores con el software libre R)

Tabla 5.18. Valores simulados x de la variable aleatoria X tiempos entre llegadas de 50 clientes en el Banco WW

u	x
0,559355	10
0,982616	18
0,325262	8
0,535503	10
0,507678	10
0,281995	8
0,347206	8
0,921085	16
0,408148	8
0,095737	6
0,333891	8
0,620771	12
0,641131	12
0,752135	12
0,204952	8

(Los valores de esta tabla fueron generados por los autores con el software libre R)

Ejercicios capítulo 5

- 5.1 Una variable aleatoria presenta una función de densidad proveniente de la Distribución Lambda Generalizada con parámetros:

$$\lambda_1(0,1) = -0.5331, \lambda_2(0,1) = 0.034, \lambda_3 = 0.0097, \lambda_4 = 0.0285$$

- i) Elaborar una tabla de valores y trazar la figura de la función de densidad. ¿Es asimétrica?, ¿es platicúrtica?
- ii) Obtener su valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- iii) Simular 50 valores de esa variable aleatoria.

- 5.2 Una variable aleatoria presenta una función de densidad proveniente de la Distribución Lambda Generalizada con parámetros:

$$\lambda_1(0,1) = 2.1, \lambda_2(0,1) = 0.1010, \lambda_3 = 0.06222, \lambda_4 = 0.06222$$

- i) Elaborar una tabla de valores y trazar la figura de la función de densidad. ¿Es simétrica?, ¿es leptocúrtica?
- ii) Obtener su valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- iii) Simular 100 valores de dicha variable aleatoria

- 5.3 Una variable aleatoria presenta una función de densidad proveniente de la Distribución Lambda Generalizada con parámetros:

$$\lambda_1(0,1) = 0.8596, \lambda_2(0,1) = 0.009543, \lambda_3 = 0.002058, \lambda_4 = 0.023$$

- i) Elaborar una tabla de valores y trazar la Figura de la función de densidad. ¿Es similar a una densidad chi cuadrado?
- ii) Obtener su valor esperado, la varianza, el coeficiente de asimetría y el de curtosis.
- iii) Simular 80 valores de dicha variable aleatoria

- 5.4 El tiempo que se necesita para lavar un vehículo en la estación de servicio MM se ajusta a una distribución uniforme con tiempo entre 10 y 15 minutos. Se requiere simular el tiempo que será necesario para lavar cada uno de los 20 vehículos que se espera lleguen en las próximas horas.

- 5.5 El tiempo de servicio en las cajas del Banco HH tiene una distribución exponencial de parámetro 9 minutos. Elaborar un proceso de simulación para los tiempos de servicio de los 30 clientes que se encuentran en la cola de dicho banco.

Referencias

- Abramowitz, M. & Stegun, I. A. (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. *National Bureau of Standards Applied Mathematics Series 55*.
- Albert, J. (2009). *Bayesian computation with R*. New York: Springer.
- Aliaga, M. y Gunderson, B. (2005). *Interactive Statistics*. México: Prentice Hall.
- Azarang, M. (1996). *Simulación y análisis de modelos estocásticos*. México: McGraw Hill.
- Barceló, J. (1996). *Simulación de sistemas discretos*. Madrid: Isdefe.
- Barros, M., Paula, G. A., & Leiva, V. (2009). An R implementation for generalize Birnbaum–Saunders distributions. *Computational Statistics & Data Analysis*, 53(4), 1511-1528.
- Batanero, C. (2001). Didáctica de la Estadística. *Granada: Universidad de Granada*.
- Bickel, P., & Doksum, K. (1977). *Mathematical Statistics: Basic ideas and select topics*. San Francisco: Holden-Day, Inc.
- Birnbaum, Z. W. and Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability* 6: 319-327.
- Birnbaum, Z. W. and Saunders, S. C. (1969b). Estimation for a family of life Distributions With Applications to Fatigue. *Journal Applied Probability* 6: 328-347.
- Blanco, L. (2004). *Probabilidad*. Bogotá: Univ. Nacional de Colombia.
- Burbano, V. M. (2010). Una manera alternativa de simular variables aleatorias con distribución normal, uniforme y logística. *Revista Ciencia en Desarrollo*, 3(1), 63-72.
- Chang, D. S. & Tang, L.C. (1994). Randon number generator for the Birnbaum-Saunders distribution. *Comput Indian Eng*, 27:345-348
- Chang, D. S., & Tang, L.C. (1994). Percentile bounds and tolerance limits for the Birnbaum-Saunders distribution. *Communications in Statistics: Theory and Methods* 23: 2853-2863.
- Churchman, C. W. (1973). *El enfoque de sistemas*. México: Diana.

- Desmond, A. F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics* 13: 171-183.
- Díaz, J. A. & Leiva, V. (2005). A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128(2), 445-457.
- Freudenthal, A. M. and Shinozuka, M. (1961). Structural Safety under Conditions of Ultimate Load Failure y Fatigue. *Technical Report. Wright Air Development Division, Wright Air Force Base*: 61-77. Dayton, OH.
- Glasserman, P. (2004). *Monte- Carlo Methods in Financial Engineering*. Computational Finance. New York: Springer Verlag.
- Gentle, J. E. (1998). *Random number generation and Monte Carlo methods*. New York: Springer.
- Godino, J. D., Batanero, C. & Cañizares, M. J. (1987). *Azar y probabilidad. Fundamentos didácticos y propuestas curriculares*. Madrid: Síntesis.
- Heitele, D. (1975). An epistemological view and fundamental stochastic ideas. *Educational studies in Mathematics*, 6(2), 187-205.
- Hill, H. C., Ball, D. L. y Schilling, S. G. (2008.) Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39, 372-400.
- Jones, G., Langrall, C y Mooney, E. (2007). Research in probability. Responding to classroom realities, en F. K. Lester (ed.), *Second Handbook of Research on Mathematics Teaching and Learning*, Charlotte, Information Age Publishing, pp. 909-955.
- Kandu, D., Kannan, N., & Balakrishnan, N. (2008). On the hazard function of the Birnbaum-Saunders distribution and associated inference. *Computational Statistics & Data Analysis* 52: 2692-2702.
- Karian, Z. A., Dudewics, E. J. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. Boca Ratón, FL: CRC Press.
- Law M. A. M. y Kelton W. D. (1991). *Simulation Modeling and Analysis*. México: McGraw Hill.
- Lehmer, D. H. (1951). Mathematical methods in large-scale computing units. *Ann. Comput. Lab. Harvard Univ*, 26, 141-146.
- Lindgren, B. (1993). *Statistical Theory*. New York: Chapman & Hall.
- Leiva, V., Sanhueza, A., & Angulo, J. M. (2009). A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment*, 23(3), 299-307.
- Marshall, A. W., Olkım, I. (2007). *Life distributions*. New York: Springer.
- Marsaglia, G. (1963). Generating Discrete Random Variables in a Computer, en *Commun Assoc. Comput. Mach.*, 6, pp. 37-38.

- Michael, J. R., Schucany, W. R. and Haas, R. W. (1976). Generating random variates using transformations with multiple roots. *The American Statistician*, 30, 88-90.
- Mayorga, J. H. (2003). *Inferencia estadística*. Bogotá: Univ. Nacional de Colombia.
- Naylor, T. (1993). *Técnicas de simulación en computadoras*. México: Limusa.
- Papoulis, A. (1991). *Probability, Random variables and Stochastic Process*. New York: McGraw-Hill Inc.
- Poincaré, H. (1936). El azar. Artículo publicado originalmente en lengua inglesa en *Journal of the American Statistical Association*, 31, 10-30.
- Polya, G. (1982). *Cómo plantear y resolver problemas*. México: Trillas.
- Ríos, I. D. (2000). *Simulación, Métodos y Aplicaciones*. Bogotá: Alfa Omega.
- Ross, S. M. (1998). *A First Course in Probability*. México: Prentice Hall.
- Ross, S. M. (1999). *Simulación*. México: Prentice Hall.
- Saavedra, P. e Ibarra, V. (2008). *Simulación numérica de variables aleatorias*. Univ. Autónoma Metropolitana-Iztapalapa.
- Saunders, S. C. (2007). *Reliability, Life Testing and Prediction of Services lives*. New York: Springer.
- Shao, J. (1999): *Mathematical Statistics*. New York: Springer.
- Turing, A. (1950). Máquinas de calcular e inteligencia [Calculate machines and intelligence]. En A. Amderson (Ed.), *Controversia sobre mentes y máquinas* (p. 56). Barcelona: Tusquets.
- Valdivieso, M. A. (2010). *Probabilidad Básica y distribuciones. Apoyo al estudio independiente*. Tunja: Impresiones Jotamar.

Este libro se terminó de imprimir en agosto de 2014, en la Imprenta de la Universidad Pedagógica y Tecnológica de Colombia.

Tunja – Boyacá – Colombia