

# 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  $\alpha$ . Este  $\alpha$  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:  $1 - \alpha = 0.95$ ;  $\alpha = 0.05$ ;  $\bar{X} = 0,478$

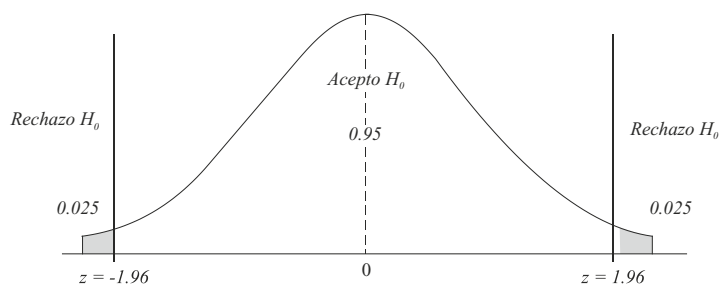
Al aplicar la prueba de Kolmogorov-Smirnov para los 40 números anteriores, se tiene que la prueba resulta no significativa, es decir, siguen una distribución uniforme de media 0.5 en el intervalo  $(0,1)$ .

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**

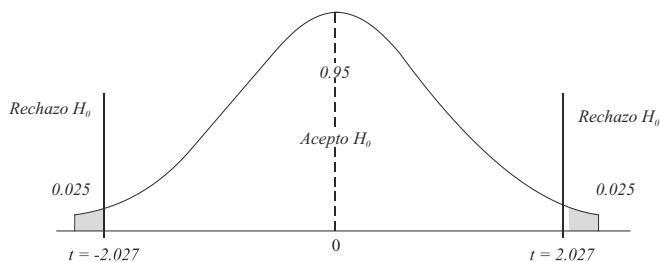
(Esta gráfica de la distribución normal estándar fue elaborada por los autores con el software libre R).

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  $\alpha$  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	+

$h = 40$

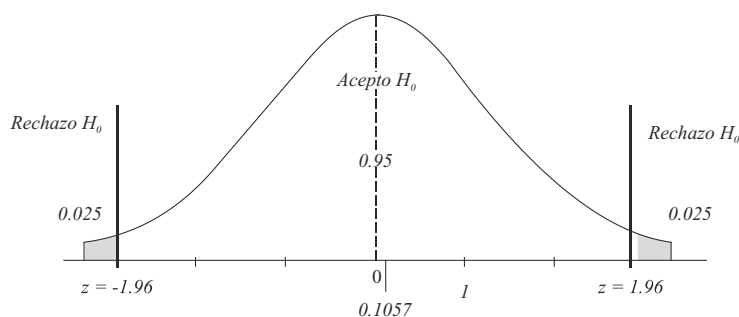
$n = 60$

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

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

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

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



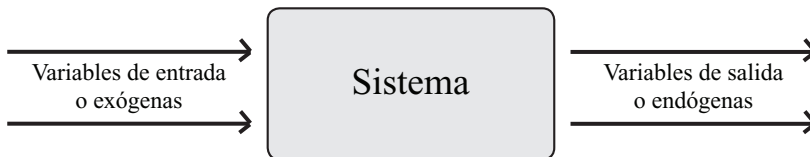
Con esos elementos se debe hacer un diagrama de bloque que puede involucrar tres tipos de variables:

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.