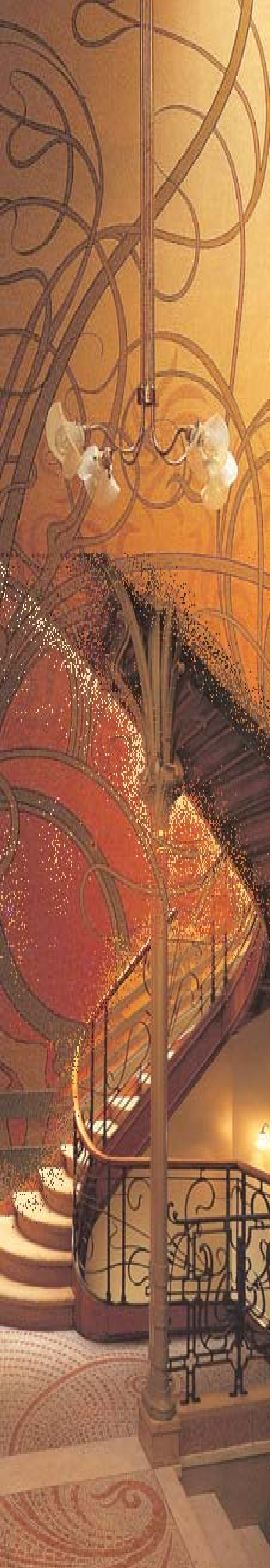
Conceptos Básicos de Estadística para Simulación

Autor:

Dra. Idalia Flores de la Mota

Octubre 2011

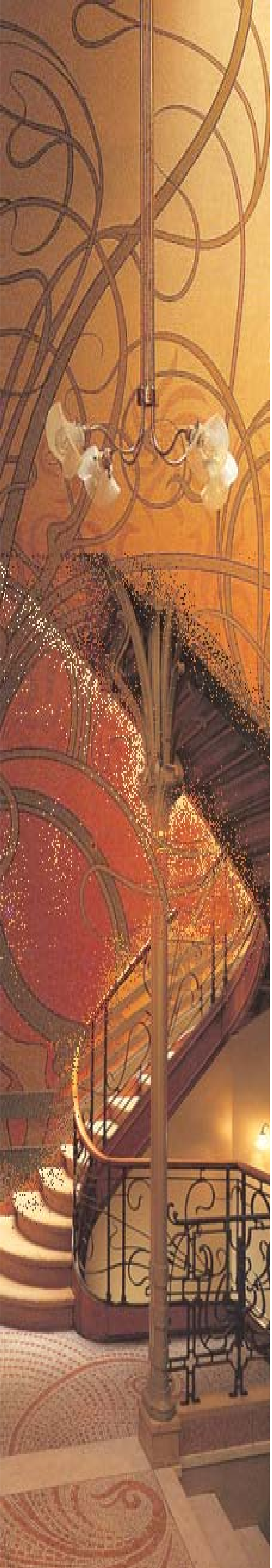


FLORES de la Mota Idalia

Conceptos Básicos de Estadística para Simulación
Universidad Nacional Autónoma de México.
Facultad de Ingeniería, 2011.

Conceptos Básicos de Estadística para Simulación

Primera edición, octubre 2011



Agradecimientos

A la Dirección General de Asuntos de Personal Académico (DGAPA) por el apoyo económico proporcionado para la elaboración de este material, a través del proyecto “Aplicaciones de la estadística en la ingeniería” PAPIME 104311. En particular a Yéssika Guzmán López, Alfonso Salazar Moreno y María Guadalupe Ávila por su apoyo tanto en las gestiones administrativas derivadas de este proyecto como en el desarrollo del material. Y a la División de Ingeniería Mecánica e Industrial por las gestiones hechas para la consecución de dicho proyecto.

Agradezco el apoyo en la revisión del material a Odette Alejandra Pliego Martínez, y a los alumnos de la materia de Simulación por la aportación de los ejemplos de las distribuciones de probabilidad.

PROLOGO

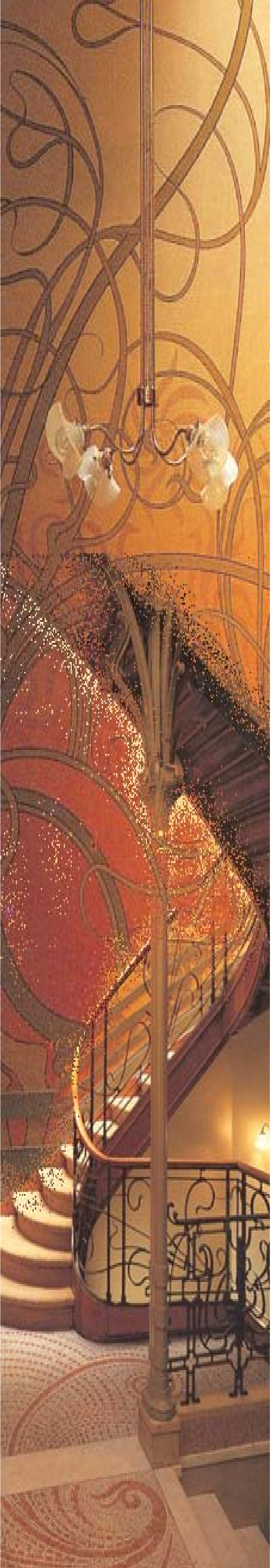
Como parte de las actividades de la Sección de Investigación de Operaciones e Ingeniería Industrial del Departamento de Sistemas de la División de Ingeniería Mecánica e Industrial, Facultad de Ingeniería UNAM, nos hemos propuesto el desarrollo de material didáctico y de divulgación a través de la elaboración de una serie de cuadernillos de difusión así como apuntes que complementen la bibliografía de los cursos de la Facultad. Este material se centra en los tópicos que abordan la estadística aplicada a la Ingeniería con el propósito, de apoyar tanto la labor docente así como el aumento en el conocimiento de la estadística por los estudiantes a nivel de maestría y licenciatura de las diferentes áreas de la ingeniería así como de licenciaturas afines, y que conozcan las aplicaciones y el potencial que ofrecen estas herramientas.

Este material tiene como uno de sus objetivos introducir al lector en los temas de una manera clara y sencilla pero sin perder el rigor teórico en su tratamiento. La selección, edición y revisión técnica del material fue realizada por un grupo de profesores del departamento de sistemas, que incluye además de la responsable del proyecto a los profesores participantes: Idalia Flores de la Mota, Víctor Damián Pinilla Morán, Francisco Álvarez Echeverría, Javier Cortés y Miguel Eduardo González Cárdenas.

Existe un proyecto PAPIME anterior ("Cuadernos de divulgación sobre aplicaciones de la ingeniería industrial y la investigación de operaciones" 105407), así como otro proyecto PAPIME vigente ("Apoyo multimedio para la aplicación de la estadística en temas de calidad" 102811) los cuales divulgan material didáctico aunque tratan temas diferentes, comparten el mismo propósito de proporcionar herramientas de apoyo a docentes y estudiantes. Por lo que los cuadernillos de divulgación generados en este esfuerzo se realizarán con numeración consecutiva y formato similar que los de los otros dos proyectos mencionados.

Fue tarea agradable la redacción de este material, y los autores esperan que los estudiantes de ingeniería lo encuentren agradable e informativo, cuando traten de aprender cómo debe aplicarse la estadística en sus campos de interés.

Francis Soler Anguiano
Responsable del proyecto



Idalia Flores de la Mota Estudió el Doctorado en Investigación de Operaciones en la Facultad de Ingeniería de la UNAM. Obtuvo mención honorífica en la Maestría y medalla Gabino Barreda por el mejor promedio de su generación. Matemática por la Facultad de Ciencias de la UNAM. Ha sido miembro de SIN. Desde 1987 forma parte del personal académico de la Facultad de Ingeniería en el departamento de Sistemas, en las categorías de: Profesor Asociado “B”; obteniendo en 1996 la categoría de profesor titular “A” Definitivo de Tiempo Completo. Nivel C de Primas de Desempeño del Personal Académico (PRIDE)

De 1992 a 1996 fue Coordinadora de la Sección de Investigación de Operaciones del Departamento de Sistemas de la DEPFI, UNAM; Retomando esta ardua labor desde el año 2000 a la fecha, ahora fungiendo como Coordinadora de las Maestrías de Ingeniería en Investigación de Operaciones e Ingeniería Industrial, y además a partir del 2005 de la Maestría en Gestión Integral del Agua. Del 2000 al 2005 Coordinó la maestría en Operación Financiera. Desde el 2001 es miembro del Subcomité Académico del Campo de Conocimiento de Sistemas. En el 2002, participó en el Plan de Desarrollo de la Facultad de Ingeniería, en lo relativo a Posgrado. En ese mismo año obtuvo su acreditación como tutora de doctorado en el Programa de Posgrado de Ingeniería, por parte del Comité Académico y miembro del Padrón de Tutores de Maestría y Doctorado del mismo programa; además, se incorpora al Subcomité de Administración del Programa de Maestría y Doctorado en Ingeniería. En el 2004 realizó la puesta en marcha de la Maestría en Ingeniería Industrial, en la Facultad de Ingeniería. Participa en la Revisión de Planes y Programas de Estudio de los posgrados que tiene a su cargo. En el 2006 realizó la planeación, coordinación y logística del “40 Aniversario de la Maestría y Doctorado en Investigación de Operaciones” con la participación de personalidades académicas nacionales e internacionales. Ha impartido maestrías en la Facultad de Ingeniería a nivel Posgrado desde 1990 de: Teoría de Redes, Programación Entera, Teoría y Técnicas de Operación I, Álgebra Lineal, Matemáticas, Probabilidad, Programación Dinámica, Simulación, Desarrollo de Habilidades Directivas, Metodología de la Investigación de Operaciones, Trabajo de Investigación I y II, Seminario de Investigación, Seminario de Investigación Doctoral. Además de impartir cursos en universidades del extranjero. Participa en Convenios de Colaboración con Sedesol, CFE, IMAT, ADO, desde el 2001 como expositora en los Cursor Regionales y en el Desarrollo de Estudios Técnicos.

Índice

INTRODUCCIÓN

En los modelos de simulación siempre se tiene como antecedente el uso de estadística ya que el carácter aleatorio de los mismos hace necesario que se haga uso de distribuciones de probabilidad. Es decir, un modelo de simulación involucra la recolección de datos para la construcción del modelo, para tal objetivo se requiere contestar algunas preguntas como: ¿Con qué información contamos? Hasta hace algunos años, el principal problema era que no existía información concentrada, había que diseñar estrategias para su obtención y sobre todo ser suficientemente creativos para buscar fuentes alternas de información. En consecuencia, un fracaso común en los estudios de simulación que no son bien delimitados en la etapa de planeación, se debe a que de la simulación se extraen más datos de los necesarios o de los que pueden validarse con los datos disponibles. Algunas preguntas que pueden apoyar este proceso son:

- ¿Qué datos son necesarios?
- ¿Cómo se obtendrán esos datos?
- ¿Qué tiempo aproximado tomará la realización de cada etapa de la obtención de datos?
- ¿Con qué información y cómo se validarán los resultados de la simulación?
- ¿Cuáles configuraciones del modelo se deberían correr?
- ¿Cuántas y qué tan grandes deben ser las corridas?

Para recolectar información de la estructura del sistema y los procedimientos de operación, es necesario hacer las siguientes consideraciones:

- No es suficiente un solo documento o la entrevista con una persona. Para el analista en simulación es fundamental hablar con tantos expertos en el sistema como sea necesario, para obtener un entendimiento completo del sistema a modelar.
- Parte de la información proporcionada será invariablemente incorrecta. Si cierta parte del sistema es particularmente importante, entonces al menos se requerirán dos expertos en el sistema.
- Los procedimientos de operación del sistema pueden no estar formalizados.

La recolección de datos (si es posible) sirve para especificar los parámetros del modelo y las *distribuciones de probabilidad* (por ejemplo para el tiempo de falla y el tiempo de reparación de la máquina). La simulación de un sistema o proceso donde hay componentes que inherentemente son aleatorios, requiere la generación de *variables aleatorias.* En las secciones siguientes se discute cómo estos valores pueden generarse conveniente y eficientemente a partir de una distribución de probabilidad deseada, para poder usarse en los modelos de simulación, se incluyen además las funciones de distribución que se usan más en modelos de simulación y se especifica en que casos se usan. Se debe tener cuidado para no caer en dos errores comunes a este nivel: reemplazar una distribución de probabilidad por su valor medio o el uso de una distribución inapropiada. Este cuadernillo tiene como finalidad mostrar algunos elementos necesarios de estadística para su uso en modelos de simulación discreta. Y está organizado como sigue, capítulo 1 en que consiste la simulación digital discreta, capítulo 2 generación de números aleatorios, capítulo 3 distribuciones de probabilidad continuas, capítulo 4 algunas distribuciones de probabilidad discretas, en estos dos capítulos (tres y cuatro) se dan ejemplos proporcionados por estudiantes del curso de Simulación de la Maestría en Investigación de Operaciones del Posgrado en Ingeniería, UNAM.

CAPÍTULO 1

SIMULACIÓN DIGITAL

1.1 Introducción

¿Qué es la *simulación?* Un concepto intuitivo es: simular, es representar, fingir, actuar. En la ciencia, la industria y la educación no es algo distinto: la simulación es una técnica de investigación o enseñanza, que reproduce en forma semejante o aproximada los eventos reales y los procesa con ciertas condiciones de prueba, definidas con anterioridad. Desarrollar simulaciones de este tipo requiere de procesos matemáticos, que en algunos casos son complejos. Inicialmente, debe especificarse un conjunto de reglas, relaciones y procedimientos operativos. La interacción de estos fenómenos crea nuevas situaciones o nuevas reglas que evolucionan al desarrollarse la simulación.

La forma de implementar la simulación va desde objetos muy sencillos como papel y lápiz, hasta sofisticadas representaciones en computadora, con sistemas interactivos de entornos casi reales.[FLORES2007]

1.2 Cuándo usar simulación

La simulación es una de las técnicas administrativas más frecuentemente usadas, y todo parece indicar que su popularidad va en aumento. Para analizar las razones de su uso, es interesante explorar las alternativas existentes a la simulación, es decir, los diferentes métodos que pueden usarse para resolver el mismo problema:

1. uso de algún otro tipo de modelo matemático de tipo analítico;
2. experimentación directa con el modelo real o con un prototipo de éste;
3. uso de la experimentación y la intuición.

Por otro lado, es más sencillo controlar condiciones experimentales en un modelo de simulación que en un sistema real. Podemos pensar en un modelo de un crucero vial, donde puedan analizarse diferentes sincronizaciones de semáforos sin afectar a los elementos reales, lo cual tendría un costo excesivo, que podría llegar hasta lo invaluable de una vida humana.

En un modelo de simulación es posible comprimir largos periodos de tiempo y analizar el comportamiento en forma inmediata. Podemos visualizar cómo será la población dentro de 30 años y si los servicios de transporte serán suficientes para satisfacerla.

Por supuesto, existen casos en los que el sistema que se quiere analizar ni siquiera existe, de modo que, definitivamente lo ideal será usar la simulación o algún método de tipo cualitativo. La simulación no reemplaza a otras formas de experimentación ni al juicio subjetivo, pero es una solución alternativa conveniente.[Flores 2007]

La experiencia y la intuición, así como el profundo conocimiento de los fenómenos, deberán ser ingredientes constantes para el éxito de los modelos de simulación. En resumen podemos decir que algunas ventajas al usar simulación son:

Ventajas:

- Incrementar el conocimiento sobre la dinámica del proceso estudiado
- Predecir el comportamiento del sistema
- Evaluar la sensibilidad de los parámetros del sistema
- Optimizar el comportamiento del sistema
- Explorar aquellas situaciones en las que la experimentación con el sistema real es peligrosa, problemática, cara o imposible.
- Entrenamiento

1.3 Modelos en simulación

Un modelo es la representación simplificada de un problema o situación real. Esquemáticamente lo podemos ver de la siguiente manera:

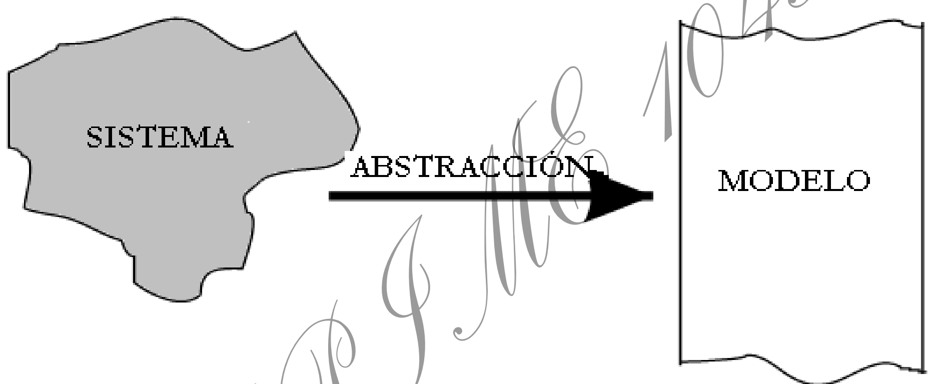


Figura 1.1 Modelación en la simulación

Esta representación puede hacerse de diferentes maneras y utilizando distintos recursos. Independientemente de cómo y con qué se haga el modelo, en cualquier caso involucra un proceso de abstracción que consiste básicamente en:

1. Seleccionar de la realidad, los elementos más importantes que intervienen en el problema y desechar aquellos que se considera que no juegan un papel determinante en el mismo.
2. Establecer con precisión las distintas relaciones que guardan entre sí dichos elementos.

Una vez realizado el proceso de abstracción se estará en condiciones de elaborar un modelo. Dependiendo de cómo y con qué se haga, tomará distintas características.

Modelo

- Un modelo es siempre una aproximación del sistema real basado en hipótesis y aproximaciones. Por tano, nunca es una representación perfecta del sistema real.
- Un modelo siempre es construido para un objetivo específico. Por lo tanto, tiene que ser formulado de forma que sea útil para responder al objetivo formulado.
- Un modelo es siempre el resultado de un compromiso entre la simplicidad y la necesidad de incorporar todos los aspectos relevantes del sistema.

Los sistemas que se tiene como objetivo modelar, pueden ser continuos o discretos en este trabajo se consideran los modelos de eventos discretos. Sin embargo y citando a Law: “Pocos sistemas en la práctica son completamente discretos o continuos, pero ya que algún tipo de cambio predomina en la mayoría de los sistemas, es siempre posible clasificarlos en discretos o continuos” [LAW 2007]. Un sistema discreto es aquél donde las variables de estado cambian solamente en un conjunto de puntos discretos en el tiempo, por ejemplo un banco es un ejemplo de un sistema discreto si consideramos como variable de estado los clientes en el banco, en cuyo caso los cambios ocurren solo cuando los clientes llegan o cuando el servicio proporcionado al cliente se ha completado, la figura 1.2 muestra como los cambios en los clientes son puntos discretos en el tiempo.[BANKS 2010]

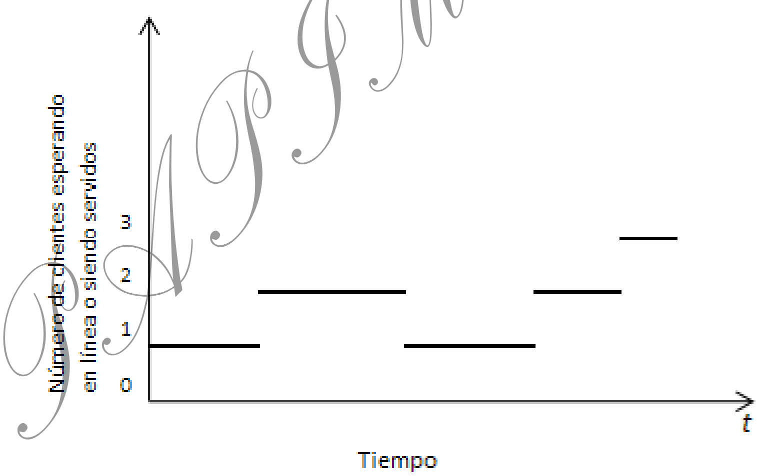


Figura 1.2 Ejemplo de sistema discreto

Un sistema continuo es aquél en el que la(s) variable(s) de estado cambia continuamente en el tiempo. Un ejemplo es el agua que esta almacenada en una presa, durante una tormenta de lluvia e incluso algún tiempo después, el agua cae en el lago detrás de la presa, posteriormente se hace fluir el agua con el propósito de controlar el flujo y producir electricidad, además de que la evaporación también hace decrecer el nivel del agua. En la figura 1.3 se muestra como la variable de estado cortina de agua detrás de la presa, cambia en este sistema continuo [BANKS 2010].

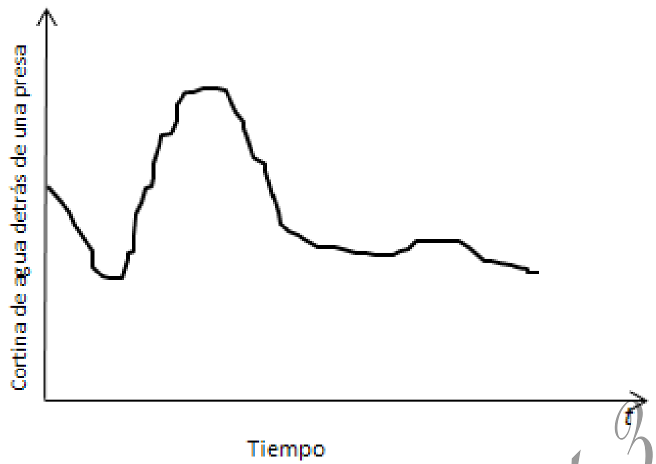


Figura 1.3 Ejemplo de sistema continuo

1.4 Ejemplo de modelo de eventos discretos [FLORES 2007]

Simulación de una barbería

El tiempo entre llegadas de clientes a "Barba Azul" tiene distribución exponencial con una media de 15 minutos. En el local hay un solo peluquero, y tarda de 10 a 15 minutos, con una distribución uniforme, para terminar un corte de pelo. A los clientes se les atiende con el sistema FIFO (first-in, first-out,"primero en llegar, primero en salir").

El objetivo de la simulación es calcular las siguientes medidas de desempeño:

- La utilización promedio del local.
- La cantidad promedio de clientes en espera.
- El tiempo de espera promedio de un cliente en la cola.

La lógica del modelo de simulación se puede escribir en términos de las acciones asociadas con sus eventos de llegada y de salida. Para lo cual se definen los siguientes eventos:

Determinación de entidades

Las entidades para este modelo con sus atributos son las siguientes:

1.	Clientes	Hora de llegada y hora de salida
2.	Barbero	Ocupado o desocupado
3.	Cola	Longitud de la cola

Determinación de eventos

Los eventos en este modelo solamente son dos y se describe a continuación la manera en que se desarrollan en la simulación.

Evento de llegada
1. Genere y guarde cronológicamente la hora de llegada del siguiente cliente = hora de simulación actual + tiempo entre llegada. Este método de generar llegadas se conoce como cordón de bota, ya que un evento de llegada genera al siguiente evento de llegada.
2. Si el recurso o instalación (el barbero) está inactivo.
   - Inicie el servicio y declare ocupado el recurso. Actualice las estadísticas de utilización del recurso.
   - Genere y guarde cronológicamente la hora de salida del cliente = hora de simulación actual + tiempo de servicio.
3. Si el recurso está ocupado, ponga al cliente en la línea de espera y actualice las estadísticas de la cola.

Evento de salida
1. Si la cola está vacía, declare inactivo el recurso. Actualice las estadísticas de utilización del recurso.
2. Si la cola no está vacía.
   - Seleccione un cliente de la cola y colóquelo en el recurso. Actualice las estadísticas de la cola y de utilización del recurso.
   - Genere y guarde cronológicamente la hora de salida del cliente = hora de simulación actual + tiempo de servicio.

Según los datos del programa, el tiempo entre llegadas tiene distribución exponencial con una media de 15 minutos, y el tiempo de servicio tiene distribución uniforme entre 10 y 15 minutos. Si p y q representan muestras aleatorias de tiempo entre llegadas y de servicio, entonces se obtiene

$$p = -15 \ln(R) \text{ minutos}, \quad 0 \leq R \leq 1$$
$$q = 10 + 5R \text{ minutos}, \quad 0 \leq R \leq 1$$

1.5 Conclusiones

Con base en todo lo que se ha dicho en este capítulo, se puede observar que dado el carácter estocástico de los modelos de simulación, es necesario contar con herramientas de probabilidad y estadística para poder desarrollar de manera eficiente tales modelos y que la aplicación de los mismos al problema real sea adecuada y con un margen de error mínimo.

Por tal motivo también es importante considerar como se trabajan los datos de inicio del modelo, si se cuenta con tales datos, si son confiables, suficientes o si en el peor caso se cuenta con pocos datos o ninguno, en este último caso se hace más necesaria la continua comunicación con los expertos del área donde se desarrolla el modelo. Todo esto se abordará con mayor detalle en los capítulos siguientes.

CAPITULO 2

GENERACIÓN DE NÚMEROS ALEATORIOS

2.1 Introducción

Los métodos de simulación aleatoria fueron inicialmente aplicados por matemáticos y físicos, para la solución de ciertos problemas determinísticos que podían expresarse en forma de ecuaciones matemáticas, cuyas soluciones no eran sencillas de obtener por los métodos numéricos o analíticos usuales. En muchos problemas matemáticos importantes, es posible encontrar un proceso estocástico, con distribución de probabilidad o parámetros que satisfagan los requerimientos de las ecuaciones. Entre los problemas determinísticos, para los cuales se ha usado la simulación estocástica, están la evaluación de integrales múltiples, solución a ecuaciones diferenciales de orden muy alto, problemas complejos de líneas de espera y programación de horarios. Aunque existen métodos analíticos para estos casos, los métodos de simulación han probado ser más efectivos.

Otro tipo de problema que conduce a la simulación de variables aleatorias surge en aquellas situaciones en las que se presenta un comportamiento estocástico, que requieren de algún tipo de muestreo, que en la práctica resulta imposible o inconveniente, como es el caso de los datos futuros. Aunque no puedan obtenerse los datos, se tiene cierto conocimiento sobre la población de la cual se originan. Para la simulación estocástica, será entonces necesario construir el modelo probabilístico adecuado al estudio. Esto significa que será indispensable identificar una distribución (o varias) de probabilidad adecuada a cada caso, la cual permita generar valores que se comporten de manera semejante al fenómeno en cuestión. Actualmente la mayoría de paquetes tanto estadísticos como de simulación incluyen un generador de números aleatorios, sin embargo en este capítulo se aborda el tema ya que es recomendable tener una idea más profunda de lo que significa generar estos números.

La metodología para generar números aleatorios, tiene una larga e interesante historia, los primeros métodos fueron prácticamente desarrollados a mano, tales como arrojar monedas, seleccionar cartas o sacar pelotas numeradas de una urna. Muchas loterías actualmente funcionan de esta manera. A principios del siglo XX, los estadísticos se unieron a los apostadores en su interés por los números aleatorios, y se construyeron mecanismos mecánicos para generar números aleatorios con mayor velocidad. En 1938 Kendall y Babington-Smith usaron un disco que giraba rápidamente para preparar una tabla de 100,000 dígitos aleatorios. Un tiempo después se desarrollaron circuitos eléctricos, basados en tubos de aspiradora, que pulsaban aleatoriamente, para arrojar números aleatorios a una tasa de 50 números por segundo. La oficina postal británica usó una máquina de este tipo, Equipo Electrónico Indicador de Números Aleatorios (ERNIE por sus siglas en inglés) para seleccionar a los ganadores de la lotería de ahorros Premium. La Rand Corporation usó otro dispositivo de este tipo para generar una tabla de un millón de dígitos aleatorios.

Se han usado muchos otros esquemas para seleccionar números aleatoriamente, como la selección de números al azar del directorio telefónico o de los reportes de censos, o usando dígitos tomados de la expansión decimal de π.

Figura 2.1 Selección de números aleatorios

En la medida en que las computadoras y también la simulación se usaron más, ha incrementado la atención a los métodos para generar números aleatorios compatibles con la manera en que trabajan las computadoras. De esta manera entre los años 40 y 50 las investigaciones se enfocaron en forma numérica o aritmética para generar números aleatorios. Dichos métodos son secuenciales, con cada nuevo número determinado por uno o varios de sus antecesores, de acuerdo con una fórmula matemática se generan nuevos, como veremos en algunos ejemplos de la siguiente sección.

2.2 Propiedades de un buen generador de números aleatorios

Debido a las características de la simulación, es necesario generar números aleatorios que representen el comportamiento del problema a simular. Aunque existen muchas formas para generar números aleatorios, para la mayoría de las aplicaciones reales, el generador debe poseer una serie de propiedades que lo hagan verdaderamente útil y similar a los procesos reales:

- Debe producir números aleatorios.
- Debe ser rápido.
- No debe requerir mucho espacio de almacenamiento en computadora.
- Debe tener un período grande antes de repetir su ciclo.
- No debe degenerar.
- Debe generar números aleatorios que puedan reproducirse.

Se explican cada una de estas propiedades de la siguiente manera:

En la producción de números aleatorios valga la redundancia, el que sean aleatorios significa que deben ser independientes entre sí. Inicialmente deben provenir de una distribución uniforme, lo que significa que en términos estrictos no son aleatorios, ya que están generados con base en una función, para efectos prácticos, es lo que más se ajusta al concepto.

Los modelos de simulación a gran escala requieren generalmente de muchos números aleatorios así, el método de generación debe ser rápido, por lo que el tiempo y la memoria utilizada en la computadora no deben ser excesivos.

Prácticamente todos los métodos de generación producen números que tarde o temprano repiten su ciclo en algún punto. Esto significa que la secuencia de números se repite. Entonces, es importante que el método de generación que se elija produzca todos los números aleatorios necesarios antes de que ocurra un ciclo completo. Esto sugiere que la selección del método estará en función de la aplicación específica. Si se requieren 100 números, un ciclo de longitud 200 no causará problemas.

También es importante que el método de generación de números aleatorios no degenere, en otras palabras, que el método no repita el mismo número en forma indefinida. Por ejemplo, algunos métodos degeneran en el valor cero.

Por tanto, se deben buscar procedimientos algorítmicos de generación de números, al menos aparentemente aleatorios. La idea de Von Neumann, era producir números que parezcan aleatorios, empleando las operaciones aritméticas de la computadora; partiendo de una semilla inicial ($u_0, u_{-1}, \ldots, u_{-p+1}$) generar una sucesión mediante $u_i = d(u_{i-1}, \ldots, u_{i-p})$ para cierta función d. Una vez elegida la semilla, la sucesión queda determinada.

La primera cuestión a resolver es qué entendemos por números aleatorios, así que es importante contar con una buena definición. Partiendo de la versión modificada de la definición clásica de Kolmogorov y Uspenskii (1987) asociada a la idea de complejidad algorítmica, tenemos la siguiente definición:

Una sucesión de números es aleatoria si no puede producirse eficientemente mediante un programa más corto que la propia serie.

Por ejemplo, la sucesión 0010010010… se interpreta como no aleatoria, dado que podemos dar un algoritmo más corto que la propia serie. La discusión de estas ideas conduce a propuestas de interés. Por ejemplo, se puede introducir un criterio de definición de números aleatorios similar al de Turing, para reconocer una inteligencia artificial, lo que nos lleva a la siguiente definición:

Una sucesión de números es aleatoria, si nadie que utilice recursos computacionales razonables en tiempo razonable, puede distinguir entre la serie y una sucesión verdaderamente aleatoria, de una forma mejor que tirando una moneda legal para decidir cuál es.

La expresión precisa de esta definición conduce a los denominados generadores PT-perfectos Lécuyer (1990), de gran interés en criptografía, pero no en simulación, debido a su lentitud.

2.3 Generación de números aleatorios con distribución uniforme entre cero y uno

Caso Discreto

La importancia de generar números aleatorios consiste en que representan el valor de una variable aleatoria, de esta manera si la variable es discreta y sólo puede tomar n valores dados que son x_i, $1 \leq i \leq n$, cuya probabilidad es p_i, sabemos que:

$$\sum_{i=1}^{n} p_i = 1$$

Estas probabilidades pueden obtenerse con anticipación, o bien determinarse mediante una serie de observaciones, a partir de las cuales se establecen las diferentes probabilidades, como se muestra en el siguiente ejemplo:

Ejemplo 2.1

En el crucero que forman las calles A y B, se realizó la siguiente observación sobre los vehículos que circulaban sobre la calle A. La tabla 2.1 muestra estas observaciones:

x_i	Observaciones	Probabilidad (p_i)	Probabilidad acumulada
a la derecha	28	0.28	0.28
a la izquierda	23	0.23	0.51
no dan vuelta	49	0.49	1.00

Tabla 2.1 Observaciones del crucero

La variable estocástica x_i puede tomar uno de los siguientes tres valores: dar vuelta a la derecha, dar vuelta a la izquierda o no dar vuelta. Observe que los posibles valores no necesariamente deben ser numéricos, pueden ser como en este caso, acciones. Lo que se exige para que sea discreta es que el número de posibles resultados sea finito con las siguientes probabilidades que se muestran en la tabla 2.2

Valor	**Derecha**	**Izquierda**	**no da vuelta**
Probabilidad	0.28	0.23	0.49

Tabla 2.2 Probabilidades asociadas al crucero

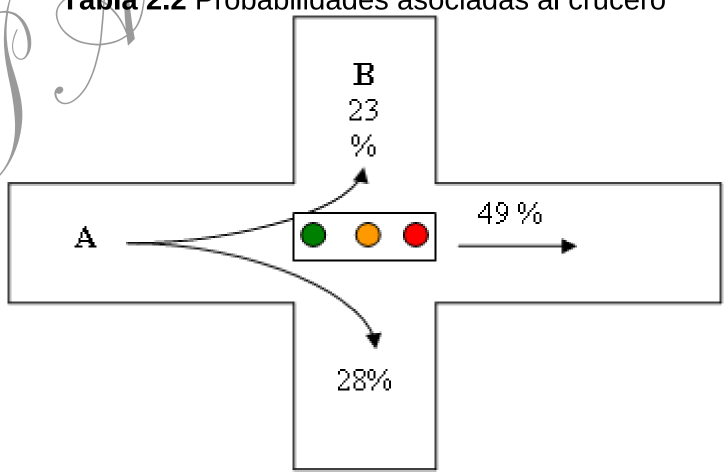


Figura 2.2 Probabilidades asociadas al crucero en forma gráfica

Para generar números aleatorios con estas probabilidades, podemos hacerlo, considerando la gráfica de la probabilidad acumulada, como la que se muestra en la figura 2.3.

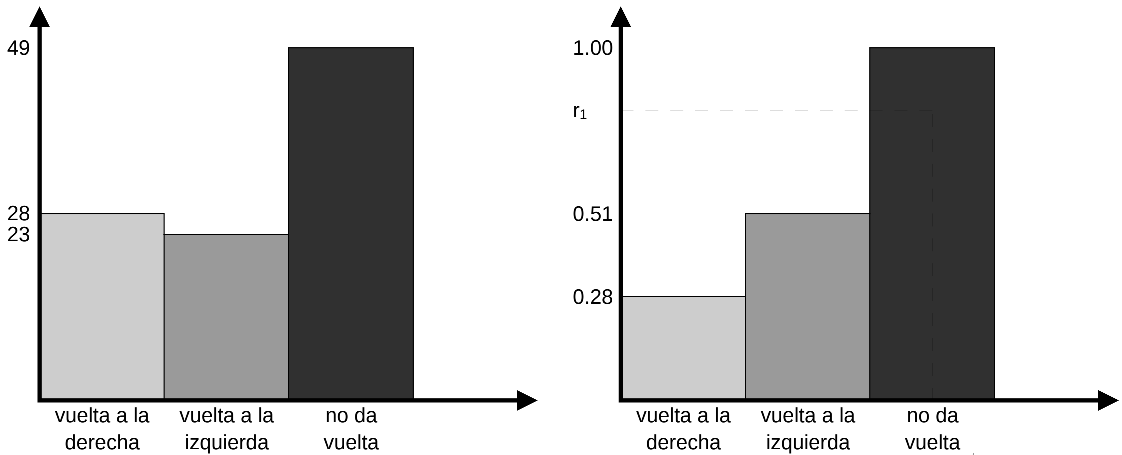


Figura 2.3 Probabilidad acumulada

Generamos una sucesión r_i, de números aleatorios con distribución uniforme entre cero y uno, y dependiendo el intervalo en el que se encuentre el número aleatorio será el valor que le asociemos, como se muestra en la tabla 2.3

r_i intervalo valor de x_i	i = 1	i = 2	i = 3	i = 4	i = 5
	0.274 (0,0.28) Derecha	0.911 (0.51,1) no da vuelta	0.046 (0, 0.28) derecha	0.466 (0.28, 0.51) izquierda	0.4976 (.0.28, 0.51) izquierda

Tabla 2.3 Intervalos y la probabilidad acumulada.

Métodos congruenciales o residuales.

Los métodos congruenciales para generar números aleatorios los sugirió por primera vez Lehmer en 1951. Estos métodos se basan en lo que en matemáticas se llama relaciones de congruencia. Aunque existen muchas variantes, los generadores congruenciales multiplicativos o métodos de potencias de residuos son los más populares.

Este método al igual que el anterior requiere de un primer número, a partir del cual se genera la serie de números aleatorios, mediante la aplicación recursiva de la siguiente fórmula:

$$X_{i+1} \equiv \alpha X_i \quad (\textit{módulo } m)$$

Esta relación se lee como "X_{i+1} es congruente con αX_i en módulo m". Por definición, se dice que dos números enteros A y B son congruentes en módulo m (con m entero), si $(A-B)$ es divisible entre m y si A y B producen residuos idénticos al dividirse entre m, esto significa que A es congruente con B en módulo m si y sólo si existe un valor k en los enteros, tal que $(A-B) = km$.

Cualquier secuencia de números puede obtenerse multiplicando el número precedente por una constante y luego reduciendo el producto por módulo *m*. La operación módulo *m* significa dividir αX_i entre *m* y conservar el residuo como el valor de X_{i+1} .

Por ejemplo, sea $\alpha = 5$, $x_o = 3$ y $m = 32$. El valor de x_1 será 15 ya que:

$$x_1 \equiv (5)\,(3)\ (\mathrm{mod}\ 32)$$
$$15/32 = 0 \text{ con 15 como residuo}$$

De igual forma puede obtenerse $x_2 = 11$. La distribución de las x_i es uniforme y es una fuente de números aleatorios.

Otra forma aún más sencilla consiste en, utilizar el random que tienen los diferentes lenguajes de programación. Es fácil ver en el ejemplo, que en el método del cuadrado del medio, la longitud máxima del ciclo de números aleatorios es de 10,000. En el caso de los congruenciales su longitud depende del módulo elegido.

CAPÍTULO 3

ELECCIÓN DE UNA FUNCIÓN DE DISTRIBUCIÓN

3.1 Introdicción

En algunos casos, el problema inicial de la simulación de variables aleatorias es precisamente la elección de una función de distribución adecuada. Pueden tomarse en cuenta cuatro consideraciones para esta selección:

1. Las características especiales de cada distribución en particular. Esto significa el comportamiento peculiar que puede tener el fenómeno en cuestión. Por ejemplo, si los datos sólo tienen dos valores distintos, es indudable que la distribución adecuada será una Bernoulli. Si se trata de datos discretos, podemos dejar de lado inmediatamente a todas las distribuciones continuas y viceversa. Si se trata de un muestreo, debemos observar si éste es con reemplazo o sin él, en este caso se usa una distribución binomial o hipergeométrica, respectivamente. Otro dato importante es la simetría o asimetría de los datos del fenómeno en cuestión, por ejemplo, la distribución normal es simétrica, mientras que una triangular puede o no serlo. Los tiempos entre eventos suelen distribuirse como exponencial, para tiempo continuo.

2. La exactitud con la que una distribución puede representar un conjunto de datos experimentales. Esto sólo se verifica a través de gráficas, como el histograma de los datos, el cual se obtuvo de las frecuencias observadas, y de pruebas de bondad de ajuste.

3. La facilidad con que la distribución se ajusta a los datos, es decir, el proceso de estimación para los parámetros correspondientes. En algunas distribuciones la obtención de estimadores es sumamente complicada y laboriosa. Sobre todo, para modelos formados por ecuaciones con parámetros no lineales. En estos casos, puede recurrirse a una distribución aproximada más sencilla o es posible hacer uso de algoritmos iterativos, para obtener estimadores suficientemente adecuados a las necesidades particulares de cada problema.

4. La eficiencia computacional para generar variables aleatorias. Como hemos ya comentado, en algunos casos se requiere de muchos cálculos para generar un conjunto de variables. En la medida en que estos cálculos sean más simples, mayor será la eficiencia de cómputo para obtener un gran número de variables, lo cual es importante, ya que se deseará contar con muestras grandes para extraer conclusiones confiables.

El empleo de probabilidad y estadística es parte integral de un estudio de simulación y se emplean para entender como modelar un sistema probabilístico que cumpla con las siguientes características:

- Para validar el modelo de simulación.
- Para escoger las distribuciones de probabilidad de entrada.
- Obtener muestras aleatorias a partir de las distribuciones.
- Efectuar análisis estadístico de los resultados de simulación.
- Diseñar los experimentos de simulación.

Se pueden tener diferentes distribuciones de probabilidad dependiendo del sistema y del problema que se quiera resolver como se muestra en la tabla siguiente:

Tipo de sistema	Fuentes de azar
Fabricación	Tiempos de proceso, tiempo entre averías, llegadas de pedidos
Comunicaciones	Tiempo entre llegadas de mensajes, duración, tipo, destino final.
Transporte	Tamaño de la carga, tiempo de transporte, tiempos de carga o descarga
Procesos hospitalarios	Tiempo entre llegadas de pacientes, tipo de enfermedad, duración de la consulta.

Tabla 3.1 Distribuciones de probabilidad de entrada.

Algunas definiciones importantes son:

- **Variable aleatoria**: **X** es una variable aleatoria si puede tomar cualquier valor de un rango finito (variable aleatoria discreta) o infinito (variable aleatoria continua). A pesar de que no se conoce la secuencia exacta de valores, si que se conoce el rango de variación y la probabilidad de que se obtenga un cierto valor.
- **Distribución de probabilidad**: permite relacionar un conjunto de valores o medidas con su frecuencia relativa de aparición.
- **Función de densidad de probabilidad ($f(x_i)$):** describe la probabilidad de que una variable aleatoria **X** tenga un cierto valor x_i.

$$f(x_i) = P(X = x_i)$$

- **Función de distribución acumulativa ($F(x_i)$):** describe la probabilidad de que una variable aleatoria **X** tenga un valor mas pequeño o igual a un cierto valor x_i.

$$F(x_i) = P(X<=x_i)$$

- **Muestreo**, es el acto de obtener muestras de una población. Una muestra debe ser representativa de la población de donde se obtiene para que sea útil al inferir estadísticamente sobre la propia población.
- **Muestra aleatoria**, Si las variables aleatorias X_1, X_2, …, X_n tienen la misma función (densidad) de probabilidad que la de la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces X_1, X_2, …, X_n forman un conjunto de n variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una muestra aleatoria de la población.

3.2 Taxonomía de las funciones de ajuste de los datos (Leemis, 2003)

En la figura siguiente se ilustra la amplia variedad de distribuciones que se pueden usar para un modelo de simulación. Los modeladores de simulación normalmente se restringen a las primeras dos ramas de la tabla, no hay una manera única de presentar esta tabla, por ejemplo las ramas que salen de *procesos estocásticos* podrían decir *estado* seguido de *tiempo*, en lugar de *tiempo* seguido de *estado*.

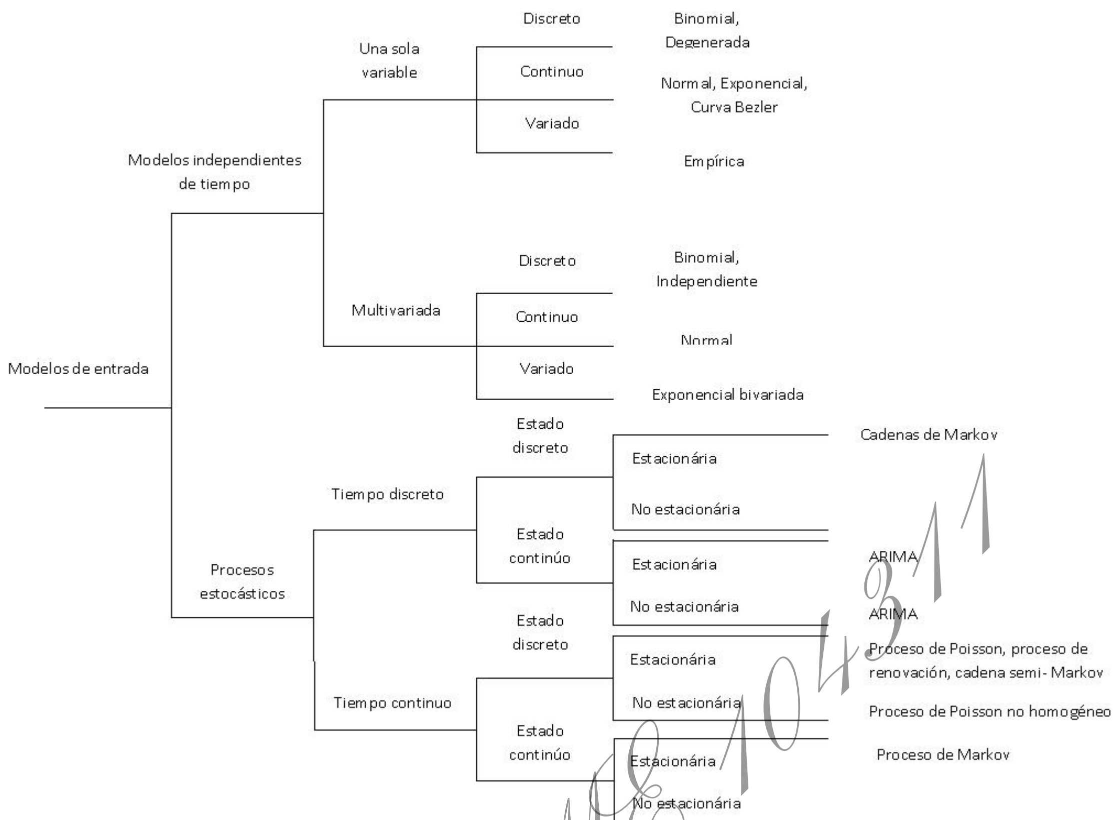


Tabla 3.2 Distribuciones de probabilidad a usar en un modelo de simulación.

Es muy importante señalar que se debe tener mucho cuidado con los datos que se recojan para efectuar una simulación ya que se tiene que verificar independencia, que sean idénticamente distribuidos y que se ajusten lo mejor posible a la distribución elegida. Actualmente los paquetes de simulación disponibles cuentan con pruebas estadísticas para el ajuste de los datos.

3.3 Funciones de distribución continua

En varios fenómenos de tipo probabilístico el comportamiento que tienen se puede asociar a funciones de distribución de probabilidad de tipo continuo. En este capítulo se presentan algunas de las funciones de distribución más comunes donde se incluye una tabla resumen, con la función de densidad de probabilidad, la función de distribución acumulativa, la media y la varianza, así como un ejemplo de la misma.

La media (o valor esperado) y la varianza de una variable aleatoria continua *X* que sigue una función de densidad *y* se calculan mediante:

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

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

3.4 Función de distribución exponencial

Esta distribución se emplea para modelar el tiempo entre llegadas. También se emplea para modelar tiempos de servicio que son muy variables, por ejemplo, la duración de una llamada telefónica. Esta distribución está relacionada con la de Poisson, dado que si una tasa de llegadas (llegadas por unidad de tiempo = λ) sigue una distribución de Poisson, el tiempo entre llegadas sigue una distribución exponencial de parámetro $\beta = 1/\lambda$.

Exponencial	$Expo(\beta)$
Posible interés	Tiempo entre llegadas de clientes cuando la frecuencia media de llegadas es constante.
Densidad de probabilidad	$f(x) = \begin{cases} \frac{1}{\beta} e^{\frac{-x}{\beta}} & x \geq 0 \\ 0 & x < 0 \end{cases}$
Distribución acumulativa	$F(x) = \begin{cases} 1 - e^{\frac{-x}{\beta}} & x \geq 0 \\ 0 & x < 0 \end{cases}$
Media	β
Varianza	β^2

Tabla 3.3 Función de distribución exponencial

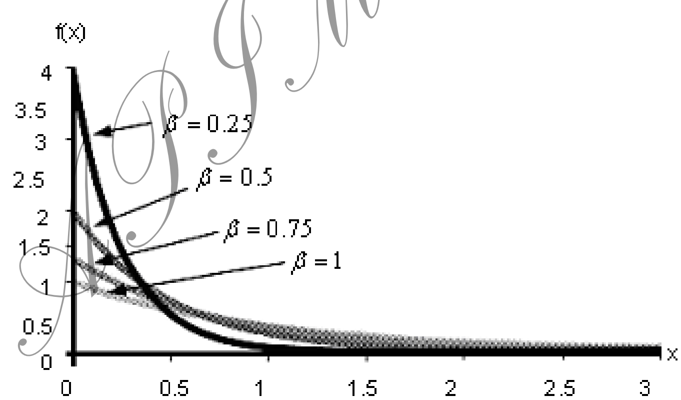


Figura 3.1 Distribución exponencial

Las funciones de distribución normal, lognormal y gamma, suelen ser las más utilizadas para modelar aquellas actividades tales que, en condiciones normales de operación, el tiempo consumido suele presentar variaciones (justificables físicamente) con respecto a un valor promedio.

Ejemplo para la función de distribución exponencial[1]

Se ha hecho un estudio sobre el tiempo de espera de los usuarios a cierto banco del DF, obteniéndose que el tiempo promedio que tardan en atender a un usuario entre las 9 y las 12 horas de un día normal es de 10 minutos, y que el tiempo de espera en ser atendido se distribuye exponencialmente.

SOLUCIÓN

Calcula la probabilidad de que si vas a dicho banco en un día normal a las 10:20 horas, seas atendido en

a) Menos de 5 minutos

Sea: **X**: Tiempo de espera en el banco hasta antes de ser atendido (en minutos).
Se sabe que **E(X)= 10 min = β**, y por la función de probabilidad acumulada:

$$P(X \leq x_i) = 1 - e^{-\frac{x_i}{\beta}}$$

Entonces, se tiene que:

$$P(X \leq 5) = 1 - e^{-0.5} = \mathbf{0.3934}$$

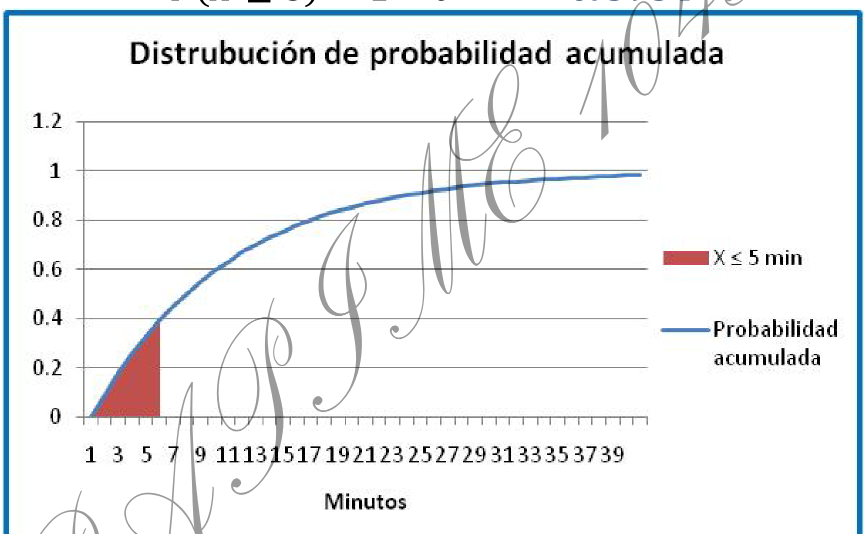


Figura 3.2 Distribución de probabilidad acumulada para x≤ 5

[1] por Jorge Andrés García Hernández, **Fuente:** Fundamentos de la teoría de las probabilidades para Ingeniería y Ciencias. Gutiérrez González, Eduardo, México, 2001, Edit. Libudi. Pag. 243

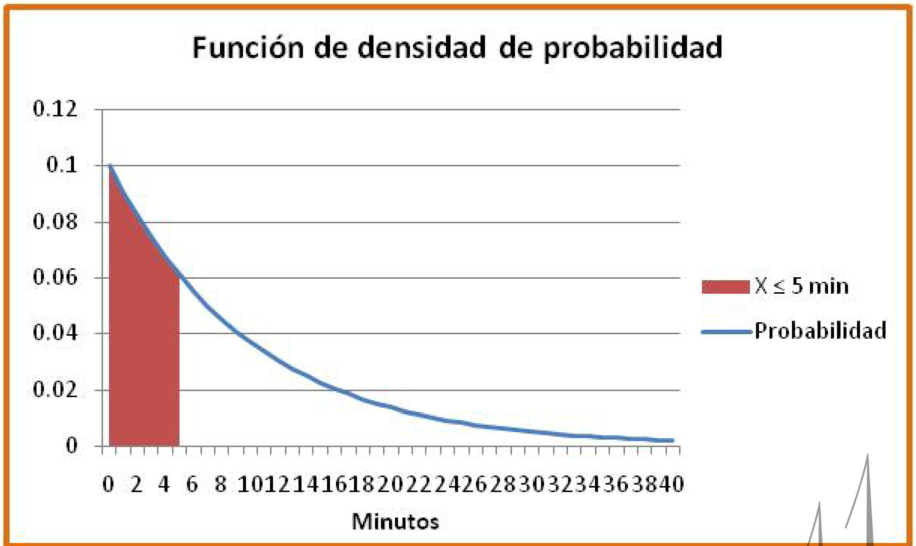


Figura 3.3 Función de densidad de probabilidad con X ≤ 5.

b) En más de 10 minutos

$$P(X \geq 10) = 1 - (1 - e^{-1}) = \mathbf{0.3679}$$

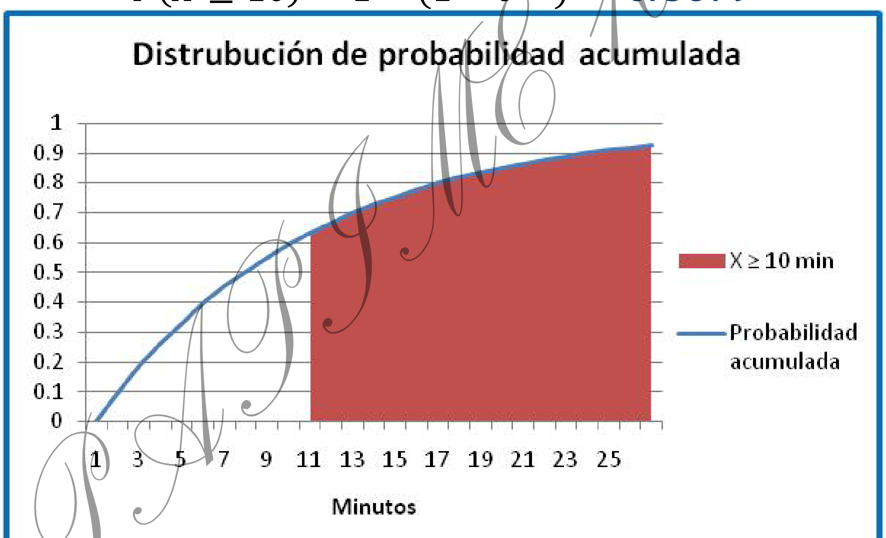


Figura 3.4 Distribución de probabilidad acumulada para X ≥ 10.

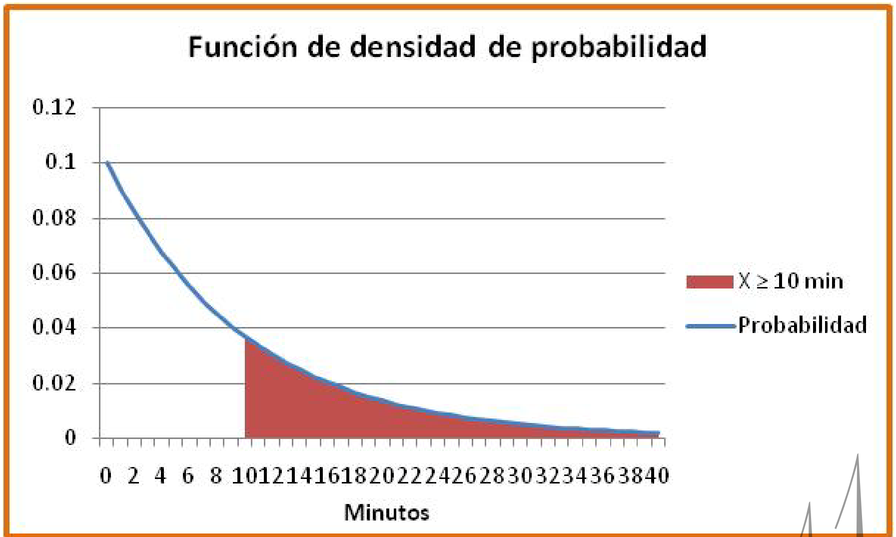


Figura 3.5 Función de densidad de probabilidad con X ≥ 10.

3.5 Función de distribución gamma

En general, el tiempo que una unidad de producción requiere para realizar una operación repetitiva de procesamiento de materia prima, o bien el tiempo consumido en una actividad repetitiva de transporte de material entre dos estaciones de trabajo, suele seguir un valor constante con pequeñas variaciones provocadas por ciertos aspectos físicos. Estos podrían ser modelados de modo determinante pero, con el objetivo de simplificar la tarea, se suelen describir como el resultado de una actividad aleatoria mediante modelos estadísticos.

En función de los parámetros de la función de distribución de probabilidad (fdp) gamma, ésta presenta una gráfica muy similar a la de la fdp normal, pero con una cierta asimetría, que responde a la presencia de datos con valores superiores al valor promedio. Esta asimetría permite modelar secuencias de actividades (por ejemplo, unidades de procesamiento o de transporte) que se realizan en paralelo, tales que cada una de ellas responde a una fdp Normal, pero el tiempo consumido en la secuencia de actividades presenta una asimetría sesgada hacia los valores superiores a la media.

Gamma	$Gamma(\alpha, \beta)$
Densidad de probabilidad	$f(x) = \begin{cases} \dfrac{\beta^{-\alpha} x^{\alpha-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$ $\Gamma(\alpha)$ es la función gamma $\Gamma(\alpha) = \int t^{\alpha-1} e^{-t} dt$ Si α es entero positivo $\Gamma(\alpha) = (\alpha - 1)!$
Distribución acumulativa	$F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} \sum_{j=0}^{\alpha-1} \dfrac{(x/\beta)^j}{j!} & x \geq 0 \\ 0 & x < 0 \end{cases}$ Si α es entero positivo; en caso contrario no hay fórmula cerrada
Media	$\alpha\beta$
Varianaza	$\alpha\beta^2$

Tabla 3.4 Función de distribución gamma

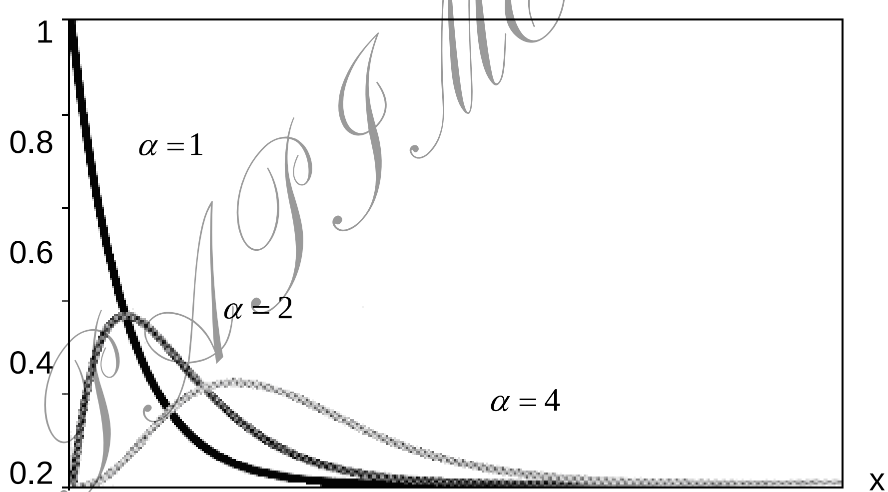


Figura 3.6 Función de distribución gamma ($\beta = 1$)

La función de distribución gamma representa una herramienta de modelado estadística muy buena para modelar sistemas reales sometidos a la aparición de ciertos eventos, por ejemplo, probabilidad de avería de una máquina, que incrementan la aparición de los valores superiores al valor promedio.

Ejemplo para la distribución gamma[2]

[2] por Jordi Messeguer Gally

En una fábrica se informa que el tiempo de reparación de una pieza tiene una distribución gamma con parámetros g(0.5,2). También se sabe que la pieza tiene un precio de venta en el mercado de $5,000 y un costo de producción de $1,000. Lo que la administración quiere conocer es el costo máximo que puede pagar por la reparación de cada pieza (dado por hora) para obtener un beneficio de $3,000.

SOLUCIÓN

Lo que se pide en el problema es un precio por hora K tal que multiplicado por el tiempo de reparación esperado nos arroje un beneficio de $3,000.

Sabemos que el beneficio o utilidad es igual al precio de venta menos la suma de los costos. Dicho en términos matemáticos para este problema:

$$\textit{Beneficio} = \textit{Precio de venta} - \textit{costo}_{\textit{producción}} - \textit{costo}_{\textit{reparación}} * \textit{tiempo}_{\textit{reparación}}$$

Sustituyendo en la ecuación anterior los datos conocidos tenemos:

$$3{,}000 = 5{,}000 - 1000 - K * E(x)$$

El valor esperado del tiempo de reparación (E(X)) es igual a la media de la distribución y se calcula dividiendo los parámetros de la función:

$$E(x) = 2/\,0.5 = 4$$

Con esto tenemos todos los datos y despejamos K. Al resolverla obtenemos que el precio que la gerencia puede pagar para obtener un beneficio de $3,000 es de $250 por hora de reparación.

3.6 Función de distribución lognormal

En general, la función de distribución lognormal se utiliza para modelar una secuencia multiplicativa (la avería de una máquina repercute en el paro del resto de máquinas) de operaciones. La función de distribución gamma se utiliza para modelar una secuencia aditiva de operaciones. La función de distribución lognormal también puede emplearse para modelar el tiempo necesario para realizar una tarea manual.

Lognormal	$LN(\mu,\sigma^2)$
Densidad de probabilidad	$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\dfrac{-(\ln x - \mu)^2}{2\sigma^2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases}$
Distribución acumulativa	No existe fórmula cerrada
Media	$e^{\mu+\sigma^2/2}$
Varianza	$e^{2\mu+\mu^2}(e^{\sigma^2}-1)$

Tabla 3.5 Función de distribución lognormal

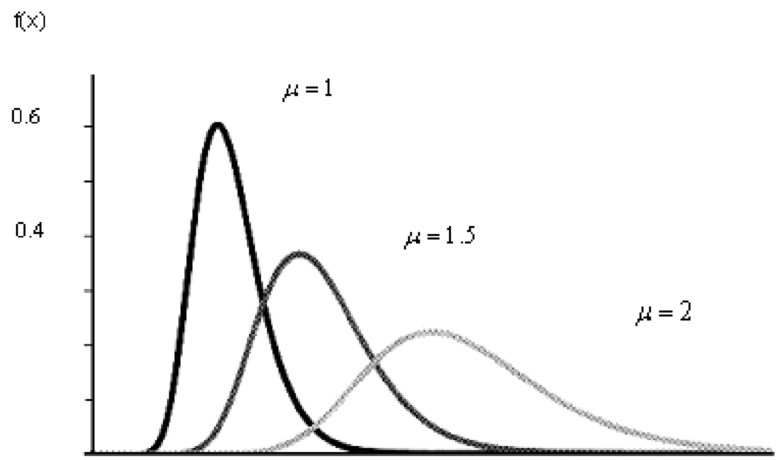


Figura 3.7 Función lognormal (σ = 1)

Ejemplo para la distribución log-normal[3]

Si W tiene una distribución lognormal, encuentre la mediana, la moda y la función de densidad de probabilidad en el punto de la moda de W, para los siguientes casos:

SOLUCIÓN:

La moda para una distribución de probabilidad lognormal está dada por $e^{\mu-\sigma^2}$ que también se puede expresar como $\exp(\mu - \sigma^2)$lo cual puede ser demostrado a partir de la definición de esta distribución y sin duda, sería una actividad interesante para el lector.

La mediana de la distribución está dada por e^{μ} , y la función de densidad de probabilidad para cualquier punto como

$$\frac{1}{x\sqrt{2\pi\sigma^2}} * e^{\frac{(lnx-\mu)^2}{2\sigma^2}}$$

Para los parámetros μ=2 y σ=0.25
Moda:

$$\frac{e^{\mu}}{e^{\sigma^2}} = e^{\mu-\sigma^2} = e^{2-(0.25)^2} = 6.9414$$

Mediana:

$$e^2 = 7.38$$

3 Por Jordi Messeguer Gally

Función de densidad de probabilidad en la moda:

$$\frac{1}{6.9414\sqrt{2\pi.25^2}}e^{\frac{(\ln(6.94)-2)^2}{2(0.25)^2}} = 0.2298e^{0.0314} = 0.0196$$

Si el lector desea repasar los conceptos se sugiere el mismo ejercicio para los siguientes parámetros: μ=1 y σ=1.5 y también μ=3 y σ=.75.

3.8 Función de distribución normal

Se utiliza para modelar sistemas donde el 70% de los datos muestreados se encuentran a una distancia inferior de σ (desviación estándar) del valor promedio μ, y la frecuencia de aparición de los datos, se encuentra distribuida simétricamente con respecto al valor promedio. Un ejemplo para utilizar una función de distribución normal es el modelado del tiempo de producción de las máquinas, cuando no se considera la posibilidad de fallos o errores de diversos tipos.

Normal	$N(\mu,\sigma^2)$
Densidad de probabilidad	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
Distribución acumulativa	No existe fórmula cerrada
Media	μ
Varianza	σ^2

Tabla 3.6 Función de distribución normal

En la figura 3.8 se ha representado el histograma de una función de distribución normal, en la cual, la diferencia de las fdp gamma y lognormal, los datos prácticamente no presentan grandes variaciones respecto a un valor promedio.

La función de distribución acumulativa no se puede calcular de forma exacta. Como consecuencia, se han empleado métodos numéricos para obtener tablas de la función. Dado que no es práctico obtener una tabla para todos los posibles valores de μ y σ^2, se construye una tabla de la distribución normal estándar (de parámetros μ = 0, σ^2=1). Si *X* es una variable aleatoria de distribución normal de valores de μ y σ^2, la variable aleatoria *Z = (X-μ)/σ* sigue una distribución normal de media 0 y varianza 1.

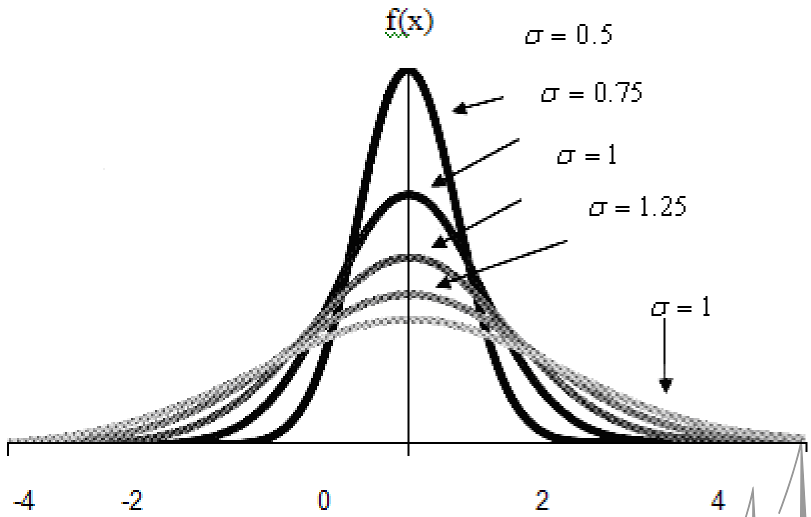


Figura 3.8 Función de distribución normal ($\mu = 0$)

Ejemplo para la distribución normal[4]
Se pretende construir un marco para la puerta de un salón de clases, en donde la estatura promedio de los estudiantes es de 5'10" con una desviación estándar de 3". Se quiere saber de qué tamaño se debe construir el marco de tal modo que el 98% de los estudiantes pueda pasar sin golpearse la cabeza.
Sabemos que para encontrar Z en una distribución normal, tenemos que:
Y según la tabla de valores para la distribución normal, el valor de Z que garantiza que quede cubierto el 98% de la población es 2.

SOLUCIÓN:
Por lo tanto, si queremos encontrar el valor de X (la altura que debería tener el marco) resolvemos:

X = 5'10" + 2(3")
X = 6'04"

La altura del marco debería ser de 6 pies con 4 pulgadas y el 98 porciento de los estudiantes lo cruzará sin problema.

[4] Por Rafael Rentería, Fuente:Introduction to probability and Statistics, autores: B.W. Lindgren, G.W. McElrath, Problema 4-26 pág 116

3.8 Función de distribución triangular

La distribución triangular proporciona una primera aproximación cuando hay poca información disponible. Esta distribución queda definida con el valor mínimo, el máximo y la moda. Se emplea para especificar actividades que tienen un tiempo mínimo, máximo y más probable.

Triangular	$Trian(a,b,c)$
Densidad de probabilidad	$f(x)\begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & c < x \le b \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x)\begin{cases} 0 & x < a \\ \dfrac{(x-a)^2}{(b-a)(c-a)} & a \le x \le c \\ 1-\dfrac{(b-x)^2}{(b-a)(b-c)} & c < x \le b \\ 1 & b < x \end{cases}$
Media	$\dfrac{a+b+c}{3}$
Varianza	$\dfrac{a^2+b^2+c^2-ab-ac-bc}{18}$

Tabla 3.7 Función de distribución triangular

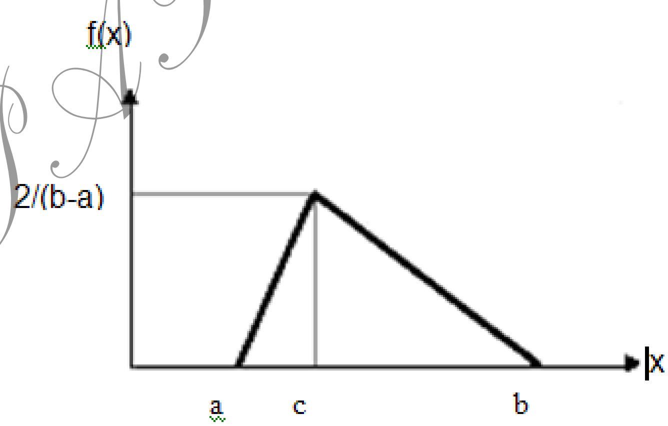


Figura 3.9. Función de distribución triangular

Ejemplo para la distribución triangular[5]

Un producto *"K"* es vendido a través de internet, dicho producto varía de acuerdo a la cantidad de demanda que se registre el día de la compra; si la demanda es mucha el costo del producto es de 50 UM , si la demanda es baja el precio es de 6 UM y si la demanda es moderada (no muy alta y no muy baja) el precio de venta es de 10 UM . Se desea conocer cuál ha sido el comportamiento de los precios del producto. Los precios, en UM correspondientes, para el producto "*K"* son:

Costo bajo	Costo medio	Costo alto
6	10	50

SOLUCIÓN:

Con base en la información anterior, podemos afirmar que:

- Los precios del producto K están entre 6 y 50 pesos, es decir, el rango= [6,50]
- En promedio el precio del producto ha sido de 22 UM, la varianza del precio es de 98.67 UM.

$$\mu = \frac{6+10+50}{3} = 22$$

$$\sigma = \frac{6^2 + 10^2 + 50^2 - (6\cdot 10) - (6\cdot 50) - (10\cdot 50)}{18} = 98.67$$

3.9 Función de distribución uniforme

La distribución uniforme es una distribución continua que se emplea para especificar una variable aleatoria, que tiene la misma probabilidad de tener su valor en cualquier punto de un rango de valores. Se define especificando un límite inferior *a* y un límite superior *b* del rango. La distribución uniforme no es, en general, una representación válida de un fenómeno aleatorio. Se emplea cuando no se conoce la distribución y sólo se tiene información de los valores extremos.

Uniforme	$U(a,b)$
Densidad de probabilidad	$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & b < x \end{cases}$

[5] por Elia Erandi González Miranda

Media	$\frac{(a+b)}{2}$
Varianza	$\frac{(b-a)^2}{12}$

Tabla 3.7. Función de distribución uniforme

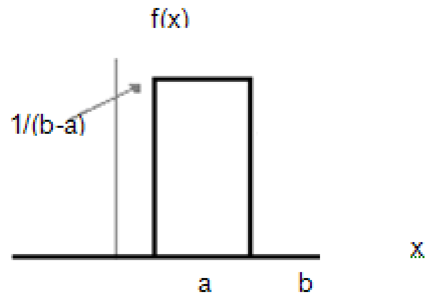


Figura 3.8 Función de distribución uniforme

Ejemplo para la distribución uniforme[6]
Un contratista A está preparando una oferta sobre un nuevo proyecto de construcción. La oferta sigue una distribución uniforme entre 55 y 75 miles de euros. Determine:

1. La probabilidad de que la oferta sea superior a 60 mil euros.
2. La media y la varianza de la oferta.

SOLUCIÓN:
Cálculo de probabilidades.
Uniforme (a,b)
a : Mínimo 55,000
b : Máximo 75,000
Punto X 60,000

Media 65,000
Varianza 33,333.33
La probabilidad de que la oferta sea superior a 60 mil euros se sitúa en un entorno de 0.75.

3.10 Función de distribución Weibull

La función de distribución Weibull es parecida a la gamma. Se emplea para modelar tiempos de proceso. También se emplea para modelar la confiabilidad de un equipo al definir el tiempo que transcurre hasta que el equipo falla.

Weibull	$Weibull(\alpha, \beta)$

[6] Por Iliana Y. Mancilla fuente: http://dxsp.sergas.es/ApliEdatos/Epidat/Ayuda/4-Ayuda%20Distribuciones%20de%20probabilidad.pdf

Densidad de probabilidad	$f(x)=\begin{cases}\alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & x\ge 0\\ 0 & x<0\end{cases}$
Distribución acumulativa	$F(x)=\begin{cases}1-e^{-(x/\beta)^{\alpha}} & x\ge 0\\ 0 & x<0\end{cases}$
Media	$\frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$
Varianza	$\frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{a}\right)-\frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$

Tabla 3.9 Función de distribución Weibull

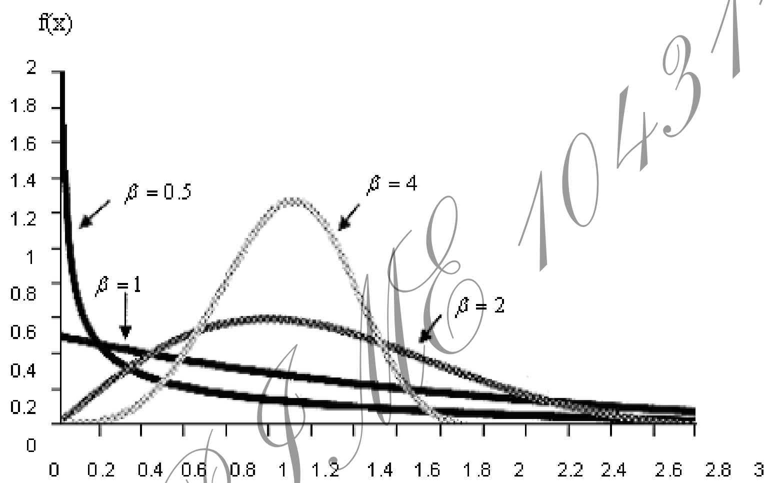


Figura 3.11 Función de distribución Weibull ($\alpha = 1/2$)

Ejemplo para la distribución de Weibull[7]

Suponga que la vida útil de una pila eléctrica para un audífono de sordera es una variable aleatoria que se distribuye Weibull con α=.5 y β=2.¿Cuánto se espera que dure la pila?

SOLUCIÓN:

$$E(x)=\frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)=\frac{2}{.5}*1!=4\ años$$

[7] Por Claudia Fabiola Peto Ramírez, Fuente: Walpole Ronald et Al. “Probabilidad y Estadística para Ingenieros”. Mc Graw Hill Interamericana de México, 1989

CAPÍTULO 4

FUNCIONES DE DISTRIBUCIÓN DISCRETA.

4.1 Introducción

Al seleccionar que distribución se debe usar para ajustar los datos de un problema que se va a simular se debe considerar si la distribución de probabilidad a usar sea discreta o continua. Si se cuenta con un número pequeño de valores se puede pensar en la conveniencia de usar una distribución discreta o si el proceso físico involucra datos discretos, como el caso de clientes a ser atendidos por un cajero.

4.2 Función de distribución de Bernoulli

La distribución de Bernoulli se aplica en los casos donde hay dos posibles estados. La probabilidad de un estado es *p* y la del otro estado *q = 1 - p*. Los fenómenos que se definen son, entre otros

- La pieza que sale de un proceso es defectuosa o no
- una empleada se presenta a trabajar o no se presenta
- una operación requiere un proceso secundario, una re-operación o no.

Bernoulli	$Bernoulli(p)$
Función probabilidad	$f(x) = \begin{cases} 1-p & si \quad x=0 \\ p & si \quad x=1 \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x) = \begin{cases} 0 & si \quad x<0 \\ 1-p & si \quad 0 \le x < 1 \\ 1 & \quad 1 \le x \end{cases}$
Media	p
Varianza	$p(1-p)$

Tabla 4.1. Función de distribución Bernoulli

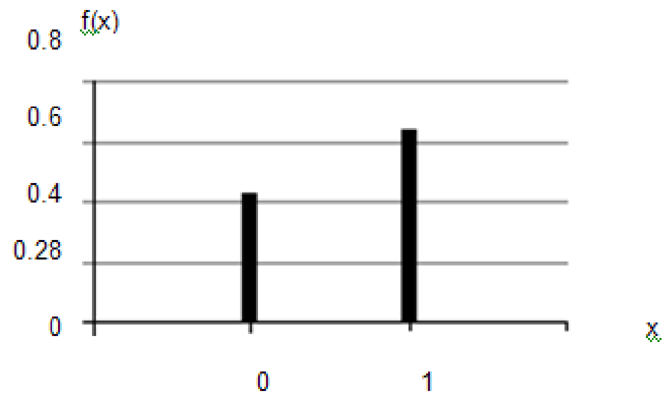


Figura 4.1 Función de probabilidad de Bernoulli (p = 0.6)

Ejemplo para la distribución Bernoulli[8]

Un estudio de tránsito muestra que de 2000 accidentes de tránsito de impacto frontal y cuyo conductor no tenía cinturón de seguridad resultaron 300 individuos con lesiones graves. Del caso anterior se puede definir a una variable aleatoria X: "Tener lesiones graves tras un accidente sin cinturón de seguridad" como una distribución de Bernoulli de parámetro (300/2000) = 0.15. Por lo tanto:

SOLUCIÓN:
p=0.15
q=(1 - p)=0.85
f(x) = 1 tiene una probabilidad de 0.15
0 tiene una probabilidad de 0.85

Esperanza Matemática
E[X]=0.15
Varianza
var[X]=p(1 - p)= pq =0.1275
Moda
0 porque q>p
La distribución tiene la siguiente gráfica:

[8] Por Israel Andrade,
Fuente:BernoulliDistribution,URL:galton.uchicago.edu/~eichler/stat22000/Handouts/l12.pdf
5 de Septiembre de 2011Estadística General, URL: www.udobasico.net/misitio/Est/DistprobII.pdf

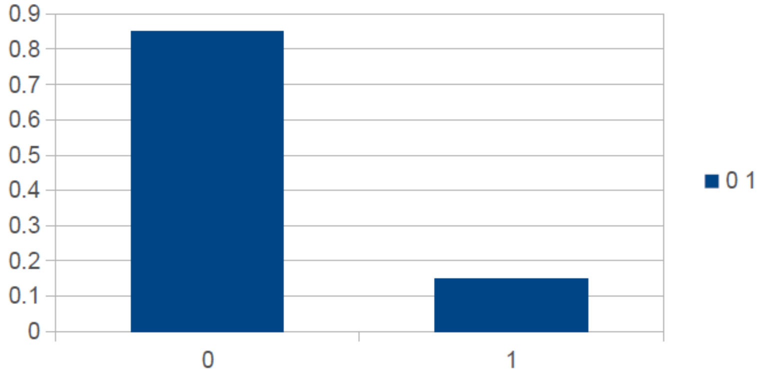


Figura 4.2 Gráfica de la distribución Bernoulli para el ejemplo.

Otros ejemplos son:

Lanzar una moneda y que salga cara: p=1/2

Elegir una persona de la población y que esté enfermo.: p=1/1000 = prevalencia de la enfermedad

Aplicar un tratamiento a un enfermo y que éste se cure: p=95%, probabilidad de que el individuo se cure.

4.3 Función de distribución uniforme discreta

Se emplea cuando todos los valores en el intervalo [*i, j*] tienen igual probabilidad. Se utiliza como un primer modelo, cuando sólo se tiene información sobre los límites del intervalo.

Uniforme discreta	$UD(i, j)$
Función probabilidad	$f(x) = \begin{cases} \dfrac{1}{j-i+1} & x \in \{i, i+1, \ldots, j\} \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x) = \begin{cases} 0 & si \quad x < i \\ \dfrac{[x]-i+1}{j-i+1} & si \quad i \le x \le j \\ 1 & j < x \end{cases}$
Media	$\dfrac{(i+j)}{2}$
Varianza	$\dfrac{(j-i+1)^2-1}{12}$

Tabla 4.2 Función de distribución uniforme discreta

Ejemplo para la distribución uniforme discreta[9]

Un turista que viaja en automóvil se encuentra extraviado en la Ciudad de Cuernavaca, su intención es ir al aeropuerto de la Ciudad de México para devolver el vehículo que rentó y abordar su vuelo de regreso a China. El problema es que no recuerda el camino de regreso y tampoco sabe hablar español.

Manejando sobre una avenida principal, llega a una ramificación de vialidades donde alcanza a leer en un señalamiento "*Hacia la Ciudad de México*". La vialidad se divide en 5 caminos, cada uno de los cuales llega hacia la Ciudad de México; uno de ellos lo lleva directamente y los demás pasan antes por diferentes ciudades. De modo que cada camino tiene un tiempo de recorrido diferente.

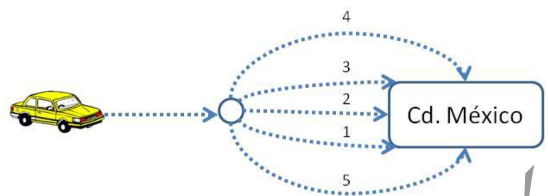


Figura 4.3 Los caminos a la ciudad de México

Debido a que el turista no puede determinar cuál de esos caminos lo lleva directamente hacia la Ciudad de México, tiene la misma probabilidad de elegir cualquiera de ellos.

En la tabla que sigue se detallan los tiempos de recorrido para cada camino:

Tabla 4.3 Tiempos para los caminos

Camino	Duración (minutos)
1	60
2	40
3	75
4	80
5	100

Con la información anterior se pide determinar:

a) El tiempo esperado de recorrido hacia la Ciudad de México.

Sea X: "Tiempo de recorrido hacia la Ciudad de México en minutos".

$$E(X) = \frac{1}{n}\sum_{k=1}^{n} x_k = \frac{1}{5(60+40+75+80+100)};$$

$$E(X) = \frac{355}{5} = 71\ minutos$$

b) La probabilidad de que el recorrido hacia México sea mayor a 75 minutos.

$$P(X > 75) = P(X = 80) + P(X = 100);$$

$$P(X > 60) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

[9] Por Jorge Andrés García Hernández

c) La probabilidad de que el recorrido tarde de 60 a 80 minutos.

$P(60 \leq X \leq 80) = P(X = 60) + P(X = 75) + P(X = 80);$

$P(60 \leq X \leq 80) = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$

d) Suponga que el combustible se consume en función del tiempo que está funcionando el motor, y que cada minuto que esté encendido, gasta $2 pesos de combustible. Considerando únicamente el tiempo hacia la Ciudad de México, **¿cuál es el gasto esperado en combustible?**

Sea:

Y: "Gasto en combustible en el trayecto hacia la Ciudad de México"

X: "Tiempo de recorrido hacia la Ciudad de México en minutos".

$Y = 2X$

$E(Y) = E(2X) = 2E(X);$

$E(Y) = 2 * 71 = \$142\ pesos$

4.4 Función de distribución binomial

La distribución binomial es una distribución discreta, que expresa el resultado de *n* experimentos independientes. Es esencialmente la suma de *n* experimentos de Bernoulli. Supongamos que un experimento que tiene dos posibles resultados se realiza *n* veces (*n* > 0). Supongamos también que la probabilidad de obtener un resultado determinado, (lo llamamos resultado ***a***), para cualquier experimento es *p*, y la probabilidad del otro resultado es *q = 1 - p* (lo llamamos resultado ***b***).

Entonces el resultado ***a***, puede aparecer un número de veces entre 0 y *n*, al igual que el resultado ***b***. La distribución binomial específica la probabilidad de que el resultado ***a*** ocurra *k* veces. Algunos fenómenos que pueden ser definidos con esta distribución son:

- El número de piezas defectuosas en un lote.
- El número de clientes de un tipo particular que entran en el sistema.

Binomial	$Bin(n, p)$
Función probabilidad	$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \{0,1,\ldots,n\} \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x) = \begin{cases} 0 & si \quad x < 0 \\ \sum_{i=0}^{[x]} \binom{n}{i} p^i (1-p)^{n-i} & si \quad i \leq x \leq j \\ 1 & \quad j < x \end{cases}$ Donde $[x]$ indica el entero más grande $\leq x$
Media	np
Varianza	$np(1-p)$

Tabla 4.4 Función de distribución binomial

Ejemplo para la distribución Binomial[10]
Se sabe que el 80% de votantes registrados en un distrito tienen más de 30 años, se quiere conocer la probabilidad que se tiene de encontrar 7 o más votantes mayores de 30 años en una muestra de 10 personas.

SOLUCIÓN:
n = 10 p (x) = nCx px qn-x
x = 7, 8, 9, 10
p = .80
q = .20

p (7 ≤ x ≤ 10) = p (7)+p (8)+p (9)+p (10)
p (7) = 10C7 (.80)7 (.20)3 = .2013
p (8) = 10C8 (.80)8 (.20)2 = .3020
p (9) = 10C9 (.80)9 (.20)1 = .2684
p (10) = 10C10 (.80)10 (.20)0 = .1074
p (7 ≤ x ≤ 10) = .2013 + .3020 + .2684 + .1074 = .8791 ≈ 87.91% de probabilidad de encontrar a siete personas mayores de 30 años en la muestra de 10.

4.5 Función de distribución de Poisson

La frecuencia de aparición de eventos en un proceso de llegadas, puede formalizarse al especificar el tiempo entre dos llegadas sucesivas, o especificando el número de eventos de llegada por intervalo.

- ***Tiempo entre 2 eventos de llegada sucesivos:*** en general, el tiempo entre dos eventos independientes de llegada consecutivos suele responder a una distribución exponencial.
- ***Número de eventos de llegada por intervalo:*** en lugar de describir el tiempo entre eventos de llegada, se describe el número de eventos en un intervalo de tiempo constante. Nótese, por ejemplo, que no es posible describir mediante una distribución exponencial la llegada de material a una unidad de producción cuando ésta es transportada en *pallets* con un número de piezas variables, ya que el tiempo entre la llegada de una pieza y la siguiente es 0. La distribución de Poisson es una de las más utilizadas para describir este tipo de comportamiento. Esta distribución fue desarrollada originalmente para modelar las llamadas telefónicas a una central. Otros fenómenos que pueden ser modelados son:

    1. El número de entidades temporales que llegan por unidad de tiempo.
    2. El número total de defectos en una pieza.
    3. El número de veces que un recurso es interrumpido por unidad de tiempo.

---
[10] Por Alexander Vindel Garduño, Fuente: Estadística para Administración y Economía, William J.Stevenson, Ed. Harla 1981

Poisson	$Poisson(\lambda)$
Función probabilidad	$f(x)=\begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x\in\{0,1,\ldots\} \\ o & resto \end{cases}$
Distribución acumulativa	$F(x)=\begin{cases} 0 & x<0 \\ e^{-\lambda}\sum_{i=0}^{[x]}\dfrac{\lambda^{i}}{i!} & o\leq x \end{cases}$
Media	λ
Varianza	λ

Tabla4.5 Función de distribución Poisson

Ejemplo para la distribución Poisson[11]

Ciertos autobuses llegan a una parada de forma aleatoria e independiente de acuerdo con un proceso Poisson. Definimos el tiempo que hay entre la llegada de un autobús y otro, como X (variable aleatoria). Se sabe que el tiempo esperado de X es de 6 minutos. Es decir, llega un autobús, en promedio, cada 6 minutos. Si yo llego al tiempo t (el cual es independiente de X), >cuánto tiempo debo esperar en promedio el autobús?:

SOLUCIÓN:
La esperanza de que el tiempo de espera sea W_t es:

$$E[W_t] = E[W_0] = 6 \text{ minutos}$$

Ya que un proceso Poisson no tiene memoria, por lo que el tiempo de espera no depende del tiempo de mi llegada.

4.6 Función de distribución geométrica

La distribución geométrica describe el número de experimentos con probabilidad *p* de éxito, que se debe efectuar hasta obtener un resultado determinado. Algunos ejemplos de fenómenos que se pueden modelar con esta distribución son:

- El número de ciclos de máquina hasta que ésta falla
- El número de piezas inspeccionadas hasta encontrar una con defectos
- El número de clientes servidos hasta que se encuentra uno de un tipo determinado.

[11] Por Citlalli Elizabeth Dorantes Bolaños, Fuente: Feller William. "Introducción a la Teoría de Probabilidades y sus Aplicaciones".Ed. Limusa. M¶exico, D.F., 1989.

Geométrica	$Geom(p)$
Función probabilidad	$f(x)=\begin{cases} p(1-p)^{x} & x\in\{0,1,\ldots\} \\ 0 & resto \end{cases}$
Distribución acumulativa	$F(x)=\begin{cases} 1-(1-p)^{[x]+1} & x\geq 0 \\ 0 & resto \end{cases}$
Media	$\frac{(1-p)}{p}$
Varianza	$\frac{(1-p)}{p^{2}}$

Tabla 4.5 Función de distribución geométrica

CAPITULO 5

ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS DE LA SIMULACIÓN

5.1 Introducción

De acuerdo a Law que en su tutorial “Como conducir un estudio exitoso de simulación” (Law, 2003), hay siete pasos para lograrlo y los vamos a analizar en este capítulo. Antes que nada, un estudio de simulación requiere que el analista que lo va a desarrollar tenga como mínimo conocimiento de la metodología de la simulación, así como bases teóricas de probabilidad y estadística, administración de proyectos, diseño y análisis de experimentos entre otros.

Algunas definiciones importantes a considerar en un estudio de simulación son (Law, 2003):

Verificación: Se refiere a poder determinar cuando un modelo de simulación conceptual (suposiciones del modelo) se ha traducido correctamente a un código de computadora o a un software comercial de simulación. Aunque el concepto de verificación es simple depurar un programa de simulación de gran escala es una tarea ardua y difícil debido al número potencialmente alto de trayectorias del programa. Algunas técnicas para depurar la simulación incluyen un recorrido estructurado del programa, uso de rastreo y depuración interactivo así como animación.

Validación: Es el proceso de determinar cuando un modelo de simulación es una representación precisa del sistema para los objetivos particulares del estudio. Si un modelo es “válido” entonces se puede usar para tomar decisiones sobre el sistema, similares a las que se podrían tomar si fuera factible y efectivo en costos en el sistema mismo.

Credibilidad es cuando un modelo de simulación y sus resultados se aceptan como “correctos” por quien toma las decisiones o administrador. La validez no implica credibilidad y viceversa. Por ejemplo, un modelo técnicamente correcto puede no ser usado en el proceso de toma de decisiones si las suposiciones clave del modelo no son claras o no se entienden por parte del que toma las decisiones. Por otra parte un modelo creíble basado en una animación impresionante de tres dimensiones puede no ser técnicamente bueno.

Datos de entrada: Es un problema estadístico que consiste en determinar que distribución de probabilidad representa mejor a los datos fuente del problema a modelar, como se vio en los capítulos 3 y 4. Las distribuciones normal o uniforme raramente son buenas para el tiempo en que se desarrolla alguna tarea.

Análisis de los datos de salida: Es un problema estadístico que se refiere a la estimación de las medidas de desempeño del modelo. Algunos temas importantes para el análisis de los datos de salida incluyen la longitud de las corridas de simulación, la longitud del período de espera (si lo hay) y el número de réplicas independientes del modelo.

5.2 Los siete pasos para conducir un estudio de simulación exitoso.(Law, 2003)

En la figura 5.1 se presentan los pasos y las actividades que se desarrollan en cada uno mismas que se analizaran en las secciones posteriores.

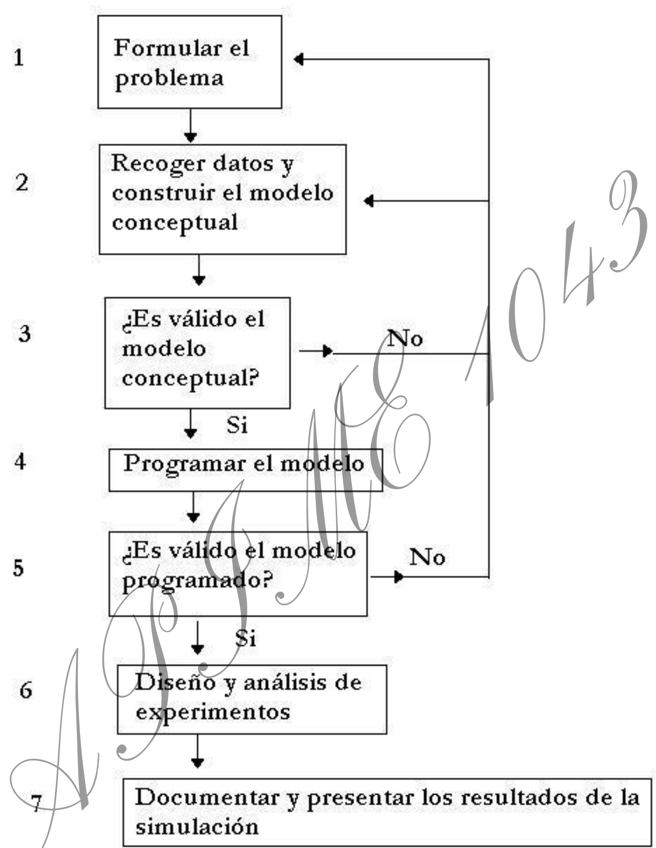


Figura 5.1 Los siete pasos

5.2 Paso 1. Formular el problema

El problema a modelar siempre es planteado por quien toma las decisiones. Note que cuando el tomador de decisiones inicia un estudio de simulación, el problema exacto a resolver muchas veces no es preciso o entendido completamente. Entonces como el estudio requiere un buen entendimiento en la medida en que se va estudiando más y delimitando el problema en cuestión queda más claro y la información que se va obteniendo debe conocerla tanto el tomador de decisiones como el equipo involucrado en la modelación de la simulación, para que de esta manera se replantee y ajuste mejor el problema a modelar.

Se deben programar reuniones al inicio del proyecto de simulación entre el administrador de proyectos, los analistas de la simulación, los expertos de área y personal involucrado que formará parte del equipo. En dichas reuniones se deben considerar los siguientes temas:

- Los objetivos generales del estudio
- Las preguntas específicas que se deben contestar en el estudio. Sin tal especificidad es imposible determinar el nivel apropiado de detalle para el modelo.
- Las medidas de desempeño que se usarán para evaluar la eficacia de las diferentes configuraciones del sistema. Diferentes medidas de desempeño algunas veces dictan diferentes niveles de detalle del modelo.
- La extensión o alcance del modelo.
- Las configuraciones del sistema que se van a modelar. Esta información es necesaria para determinar la generalidad que se debe tener en el programa de cómputo.
- El marco de tiempo requerido para el estudio así como los recursos requeridos. Los proyectos de simulación generalmente toman más tiempo del que inicialmente se estima ya que lógica del sistema se va tornando más compleja que lo que se pensaba y otra causa son los retrasos en reunir la información requerida.

5.3 Paso 2. Recabar la información o datos y construir el modelo conceptual

- Recolección de la información de la estructura del sistema y los procedimientos operativos.
    - ✓ Un solo documento o persona no es suficiente, por lo que es necesario para los analistas de simulación hablar con diferentes expertos para tener un mejor entendimiento del sistema a modelar.
    - ✓ Parte de la información proporcionada por los expertos puede ser incorrecta por lo que si dicha información es muy relevante será necesario consultar a más de un experto.
    - ✓ Puede suceder que los procedimientos operativos del sistema no estén formalizados.

- La recolección de los datos para especificar los parámetros del modelo y las distribuciones de probabilidad (por ejemplo para el tiempo de falla y reparación de una máquina). Dos errores que no deben cometerse es reemplazar la distribución de probabilidad por su media o usar una distribución inadecuada.

- Documentar las suposiciones del modelo, algoritmos y resumen de los datos en el modelo conceptual escrito (documento de hipótesis). Esta actividad es sumamente crítica que a menudo se omite –la comunicación verbal se presta mucho a errores. El modelo conceptual debe incluir:
  - ✓ Una sección de la visión general del problema que contenga las metas generales del proyecto, los problemas específicos a tratar, y las medidas de desempeño de interés.
  - ✓ Un diagrama de flujo de procesos o un diagrama de la disposición del sistema si es apropiado.
  - ✓ Descripciones detalladas de cada subsistema y cómo interactúan.
  - ✓ ¿Qué suposiciones simplificadoras se hicieron y por qué? Una simulación es una abstracción del problema real y por eso es que se simplifica de tal manera que solamente responda a preguntas de interés.
  - ✓ Resúmenes de los datos de entrada del modelo.
  - ✓ Fuentes de información importante o controversial
- Recolectar los datos de desempeño del sistema existente, si se cuenta con ellos para usarlos en el paso 5 que consiste en la validación del modelo.
- El nivel de detalle del modelo debe depender de:
  - ✓ Los objetivos del proyecto.
  - ✓ Las medidas de desempeño de interés.
  - ✓ La disponibilidad de los datos.
  - ✓ Aspectos de credibilidad.
  - ✓ Restricciones computacionales.
  - ✓ Opiniones de los expertos.
  - ✓ Restricciones de dinero y tiempo.
- No debe existir una correspondencia uno a uno entre cada elemento del modelo y cada elemento del sistema, comience con un modelo simple y vaya haciéndolo más complejo en la medida en que se vaya requiriendo. Los detalles innecesarios del modelo pueden tener como desventaja un alto tiempo de ejecución o pérdida de claridad en factores del sistema que son importantes.
- Interactuar con el que toma las decisiones tiene muchos beneficios ya que ayuda a la correcta solución del problema así como en la credibilidad del modelo al estar de acuerdo esta persona en la estructura y suposiciones del modelo.

5.4 Paso 3. ¿Es válido el modelo conceptual?

- El desarrollo de un ensayo estructurado del modelo conceptual antes de presentarlo a los interesados es una actividad importante que a menudo se omite, pero que ayuda a:
  - ✓ Asegurar que las suposiciones del modelo son correctas y están completas.
  - ✓ Fomenta la interacción entre los miembros del proyecto, ya que es necesario que cada uno de ellos lea el modelo conceptual pero no es suficiente, por lo cual si es importante que se reúnan a discutirlo.

  - ✓ Promueve la posesión del modelo lo que evitará problemas políticos en el futuro.
  - ✓ Al hacerlo antes de la programación del modelo se evita que se reprograme más tarde.
- Si se descubren errores u omisiones en el modelo conceptual, lo que siempre ocurre, entonces el modelo conceptual se actualiza antes de continuar con el paso 4.

5.5 Paso 4. Programar el modelo

Una vez que se ha elaborado el modelo conceptual se realiza la programación del mismo que puede ser usando un lenguaje de propósito general como C o C++, o usando un software comercial como Arena, Promodel o Simio entre otros. Ambos caminos tienen sus ventajas y sus desventajas, por ejemplo si se usa un lenguaje de programación se tiene mayor control sobre el programa, menor costo en cuanto al software y sobre todo que es como un "traje hecho a la medida". Por otro lado el uso de un software de simulación reduce el tiempo de programación por lo que puede reducir costos en el proyecto y cuenta con plataformas de animación que lo hacen muy atractivo visualmente, además si se quiere usar optimización la mayoría de estos productos ya cuentan con un módulo de optimización.

5.6 Paso 5. ¿Es válido el modelo programado?

Para validar el modelo ya programado se puede hacer uso de datos históricos si se cuenta con ellos, esta es la técnica de validación más importante, si se tiene éxito en la validación del modelo entonces también se gana en credibilidad del mismo. Si el sistema que se está modelando es nuevo o no se cuenta con datos históricos entonces se pueden buscar otras técnicas para validar el modelo, como optimización si es que no resulta muy complicada.

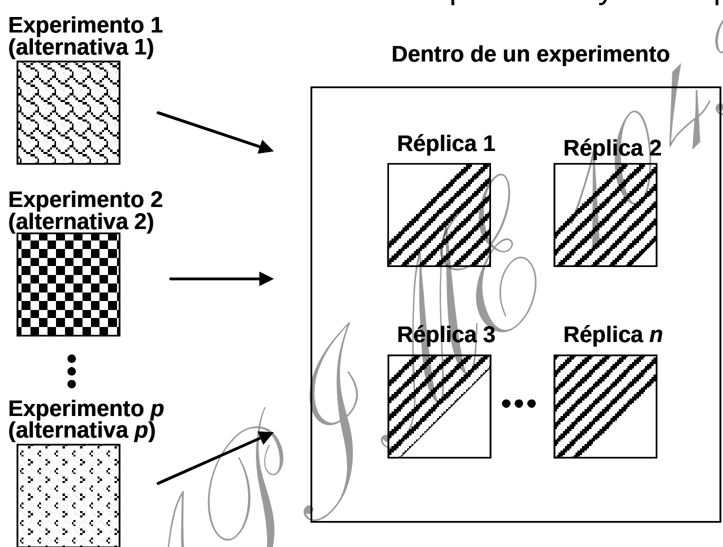
En este paso es importante que tanto los expertos como el equipo encargado del proyecto de simulación revisen los resultados para verificar que sean razonables y por lo tanto que el modelo de simulación sea válido.

Se recomienda hacer un análisis de sensibilidad del modelo programado para revisar cuales factores tienen más impacto en las medidas de desempeño y entonces tener cuidado en esa parte.

5.7 Paso 6. Diseño y Análisis de los datos de salida.

Una vez que la simulación se ha realizado es importante que se haga un análisis estadístico de los resultados de la simulación. En muchos estudios de simulación se gasta demasiado tiempo y dinero en el desarrollo del modelo y la programación, y se hace poco esfuerzo para analizar los resultados en forma adecuada. De hecho, un modo muy común de operación es hacer una sola corrida de la simulación de una longitud un tanto arbitraria y entonces usar los resultados de dicha corrida como verdaderos, sin embargo, debido a la aleatoriedad de las variables puede haber grandes variaciones en los resultados, que tienen como consecuencia inferencias erróneas del problema real. Por esto es necesario el desarrollo de uno o más experimentos, y en cada uno de ellos una o más réplicas, este patrón lo podemos ver esquemáticamente en la figura 5.1

Figura 5.1 Experimentos y réplicas en un modelo de simulación



La figura 5.1 ilustra la diferencia entre un experimento y una réplica, como se puede apreciar, las réplicas están dentro de un mismo experimento. Cuando se conduce un estudio de simulación se considera una serie de parámetros, por ejemplo, el tiempo entre llegadas. En este sentido se hace un cambio en el tiempo entre llegadas o también se puede considerar una variación en el número de servidores, el hacer cada uno de estos cambios corresponde a desarrollar diferentes experimentos para evaluar diferentes escenarios. Las réplicas son corridas del mismo modelo sin hacer cambios en los parámetros, estas réplicas arrojarán distintos resultados debido al uso de diferentes series de números aleatorios. Cada réplica produce resultados estadísticos que difieren de los producidos por otras réplicas, y dichos resultados se pueden analizar a través de todo el conjunto de réplicas.

Lo que nos interesa en esta parte del experimento es poder dar respuesta a las preguntas siguientes:

1. ¿Cuántos experimentos es necesario realizar?
2. ¿Cuántas réplicas son necesarias para la simulación?

La naturaleza aleatoria de la simulación conlleva un proceso estocástico en los resultados a los cuales podemos llamar Y_1, Y_2,…Y_m de una corrida simple. Por ejemplo Y_i podría ser el retraso en la llegada del i-ésimo trabajo en un sistema de cola simple. Se define también C_i como el costo de operar un inventario en el i-ésimo mes. Las Y_{ij} son variables aleatorias que en general no son independientes o idénticamente distribuidas (IID). De esta manera muchas de las fórmulas de la estadística clásica no serán directamente aplicables al análisis de los datos de salida de la simulación.

Ejemplo 5.1

Se considera un sistema de colas donde y_1 es el retraso del cliente 1, y_2 es el retraso del cliente 2, etc. En este sistema los retrasos en la cola no serán independientes, ya que un retraso grande de un cliente esperando en cola tenderá a retrasar aún más al siguiente cliente que espera. Suponga que la simulación inicia en el tiempo cero sin clientes en el sistema, como se hace usualmente. Entonces los retrasos en la cola al inicio de la simulación tenderán a ser menores que los que habrá al final, razón por la cual los retrasos no están idénticamente distribuidos.

Sean y_{11}, y_{12},.., y_{1m} resultados de correr la simulación para las variables aleatorias y_1, y_2,…,y_m con los números aleatorios específicos u_{11}, u_{12}, …. Si se corre la simulación con un conjunto diferente de números aleatorios u_{21}, u_{22}, …, entonces se obtiene un conjunto de resultados diferentes y_{21}, y_{22},.., y_{2m} de las mismas variables aleatorias . y_1, y_2,…y_m.

Los dos conjuntos de resultados no son iguales, ya que se usaron diferentes números aleatorios en dos corridas y por tanto se producen dos muestras diferentes de las mismas distribuciones de probabilidad.

En general suponga que se hacen n réplicas independientes o corridas de la simulación cada una de tamaño m, que en el ejemplo significaría simular los retrasos de m clientes, obteniéndose las observaciones siguientes:

$$
\begin{array}{c}
y_{11},\ldots, y_{1i},..,y_{1m} \\
y_{21},\ldots, y_{2i},..,\ _{2m} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
y_{j1},\ldots, y_{ji},...,y_{jm} \\
\ldots\ldots\ldots\ldots\ldots\ldots \\
y_{n1},\ldots, y_{ni},...,y_{nm}
\end{array}
$$

Donde y_{ji} es el retraso del cliente i en la réplica j para i = 1,…, m y j = 1,…, n .

Observación 1

Aunque se usa un conjunto de números aleatorios diferente en cada réplica, cada réplica usa las mismas condiciones iniciales y los contadores estadísticos para la simulación se reinician cada que se comienza con una nueva réplica.

Observación 2
Las observaciones de una réplica en particular (renglón) claramente no son IID. Sin embargo, note que y_{1i}, y_{2i},.., y_{ni} de la i-ésima columna, son observaciones IID de la variable aleatoria y_i, para i = 1,2, …, m.

En forma general podemos afirmar que las réplicas son independientes entre sí, y que las observaciones de cada réplica tienen la misma distribución conjunta.

Esta independencia a través de las corridas es la clave para simplificar el análisis de datos de la simulación. Por tanto hablando en términos generales el propósito del análisis de datos de los resultados de la simulación es usar las observaciones y_{ji} (i= 1,2,…,m; j= 1,2,…,n) para inferir características de las variables aleatorias y_1, y_2,…,y_m.

5.8 Documentación y presentación de los resultados de la simulación

La documentación para el modelo y el estudio de la simulación deben incluir: el modelo conceptual, que es muy importante ya que se puede volver a usar en el futuro, una descripción detallada del programa de cómputo y los resultados y conclusiones del estudio.

La presentación final de la simulación debe incluir animaciones y una discusión del proceso de construcción y validación del modelo para que con esto se pueda promover la credibilidad del modelo.

5.9 Conclusiones

Finalmente quisiera cerrar este capítulo y en general el cuadernillo explicando los siguientes aspectos:

- No se presenta en detalle el uso de herramientas estadísticas al hacer un modelo de simulación porque el propósito de este material es dar una idea general a los lectores no especializados en el tema.
- Los conceptos y resultados tanto de probabilidad y estadística pueden tener un nivel de profundidad muy alto al hacer un modelo de simulación a nivel práctico y aún más cuando es una investigación de corte teórico.
- Cuando se hace uso de software comercial para hacer un modelo de simulación se obvian muchas suposiciones teóricas de la estadística así que se debe tener mucho cuidado al hacer un ajuste de datos a una distribución, o al hacer uso de los generadores de números aleatorios, en esos casos se

pueden usar pruebas de hipótesis o algunos de los conceptos que ayuden a tener mayor certeza de que se hace un buen ajuste y no tener malas sorpresas más tarde.

- Los conceptos de estadística son una herramienta muy importante en el proceso de elaboración de un modelo de simulación, desde el inicio hasta el final por esto es importante que se tenga presente que un buen modelo no es aquél que solamente se crea en un software comercial con mucha animación pero poco respaldo teórico.

Referencias

Banks, J. 1998. (editor).: *Handbook of simulation*, John Wiley & Sons, Inc.

Barton, R. 2004.“Designing simulation experiments”, Proceedings of the 2004 Winter Simulation Conference, p.p. 73-79.

Carson, J. Banks, J. 1993.*Discrete- Event System Simulation*, Prentice Hall.

Coss Bu, R. 2003. Simulación, Un enfoque práctico, Limusa.

Feller William. 1989. *Introducción a la Teoría de Probabilidades y sus Aplicaciones*. Ed. Limusa. México, D.F.

Flores I, Elizondo M, Apuntes de Simulación, Facultad de Ingeniería UNAM, 2007.

González, M.C. 1996. Modelos y Simulación, ENEP Acatlán, UNAM.

Gutiérrez González, Eduardo, *Fundamentos de la teoría de las probabilidades para Ingeniería y Ciencias.* México, 2001, Edit. Libudi. Pag. 243

Gordon, G. 1991: *Simulación de sistemas*, 6ta reimpresión de la 1ª. Edición, Diana.

Kelton, D. Barton, R. 2003. “Experimental design for simulation”, Proceedings of the 2003 Winter Simulation Conference, p.p. 59-65.

Law, A. 2003. “How to conduct a successful simulation study”, Proceedings of the 2003 Winter Simulation Conference, p.p. 66-70.

Law, A. 2004. “Statistical analysis of simulation output data: The practical state of the art”, Proceedings of the 2004 Winter Simulation Conference, p.p. 67-72.

Law, A., KELTON, D. 2000. *Simulation Modelling and Analysis*, Mc. Graw-Hill.

Leemis 2003, “Input modeling” Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds, p.p.14-24.

Lindgren, B.W. G.W. Mc. Elrath, 1966. *Introduction to probability and Statistics*, Problema 4-26 pág 116

Stevenson ,William J., 1981 *Estadística para Administración y Economía*, Ed. Harla.

Taha, H. 2004. *Investigación de Operaciones*, 7ª. Edición, Prentice-Hall.

Walpole Ronald et Al. 1989. *Probabilidad y Estadística para Ingenieros*, Mc Graw Hill Interamericana de México.

Sitios de internet

http://dxsp.sergas.es/ApliEdatos/Epidat/Ayuda/4Ayuda%20Distribuciones%20de%20probabilidad.pdf

http://members.libreopinion.com/ve/efrain-muretti/simulacion.htm

www.udobasico.net/misitio/Est/DistprobII.pdf.