

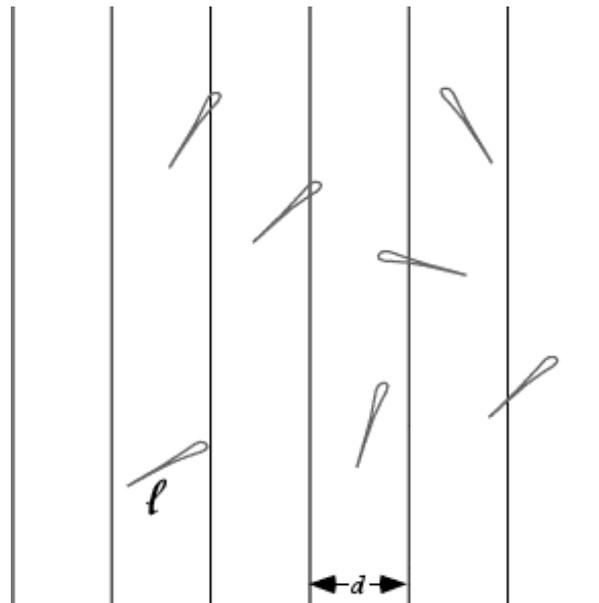


Universidad
Nacional de
Educación a
Distancia

SIMULACIÓN

Texto Base de Teoría

Curso 2008-09



Alfonso Urquía Moraleda

Departamento de Informática y Automática

Escuela Técnica Superior de Ingeniería Informática, UNED

Juan del Rosal 16, 28040 Madrid, España

E-mail: aurquia@dia.uned.es

Índice

| | | |
|-----------|---|-----------|
| I | Introducción al modelado y la simulación | 1 |
| 1. | Conceptos básicos | 3 |
| 1.1. | Sistemas y modelos | 3 |
| 1.2. | Tipos de modelos | 5 |
| 1.3. | Modelos matemáticos | 6 |
| 1.4. | El marco experimental | 7 |
| | Ejercicios | 8 |
| 2. | Pasos en un estudio de simulación | 9 |
| 2.1. | Introducción | 9 |
| 2.2. | Definición del problema y planificación del proyecto | 10 |
| 2.3. | Definición del sistema y formulación del modelo | 11 |
| 2.4. | Diseño de los experimentos | 14 |
| 2.5. | Datos de entrada | 14 |
| 2.6. | Traducción del modelo | 15 |
| 2.7. | Verificación y validación del modelo | 15 |
| 2.8. | Experimentación y análisis de los resultados | 17 |
| 2.9. | Documentación e implementación | 17 |
| | Ejercicios | 17 |
| II | Modelado y simulación | 19 |
| 3. | Método de Monte Carlo | 21 |
| 3.1. | Introducción | 21 |
| 3.2. | Estimación del valor del número π | 22 |
| 3.3. | Probabilidad de ganar un juego “solitario” de cartas | 25 |
| 3.4. | Cálculo de integrales definidas | 26 |
| | ANEXO. Programa de simulación para la estimación del valor de una integral definida | 28 |
| 4. | Modelado de sistemas de eventos discretos | 31 |
| 4.1. | Introducción | 31 |
| 4.2. | Estructura lógica del modelo | 32 |

| | | |
|------------|---|------------|
| 4.3. | Elementos del modelo | 32 |
| 4.4. | Aspectos matemáticos del modelo y las variables de entrada | 35 |
| 4.5. | Medidas del comportamiento | 36 |
| 4.6. | Condiciones inicial y final de la simulación | 37 |
| 4.7. | Descripción del funcionamiento del sistema | 38 |
| | Ejercicios | 40 |
| 5. | Simulación usando un lenguaje de programación | 41 |
| 5.1. | Introducción | 41 |
| 5.2. | Algoritmo de simulación | 42 |
| 5.3. | Simulación de una cola atendida por un empleado | 44 |
| 5.4. | Gestión del inventario de un único producto | 54 |
| 5.5. | Modelado mediante grafos de eventos | 63 |
| | Ejercicios | 66 |
| | ANEXO 1. Programa de simulación de una cola atendida por un único empleado . . | 69 |
| | ANEXO 2. Programa de simulación de la gestión del inventario de un producto . . . | 75 |
| 6. | Simulación usando Arena | 81 |
| 6.1. | Introducción | 81 |
| 6.2. | Modelado de operaciones básicas | 83 |
| 6.3. | Modelado detallado | 90 |
| | Ejercicios | 94 |
| 7. | Modelos analíticos y simulación | 101 |
| 7.1. | Introducción | 101 |
| 7.2. | Modelos analíticos frente a modelos de simulación | 102 |
| 7.3. | Cadenas de Markov | 104 |
| 7.4. | Modelos analíticos de sistemas de colas | 108 |
| III | Modelado y generación de las entradas aleatorias | 115 |
| 8. | Modelado de las entradas | 117 |
| 8.1. | Introducción | 117 |
| 8.2. | Distribuciones de probabilidad | 120 |
| 8.3. | Independencia y homogeneidad de los datos | 123 |
| 8.4. | TAREA I: Selección de la familia de distribuciones | 124 |
| 8.5. | TAREA II: Estimación de los parámetros | 130 |
| 8.6. | TAREA III: Medida de la bondad del ajuste | 136 |
| 8.7. | Selección de la distribución en ausencia de datos | 140 |
| 8.8. | Procesos estocásticos de Poisson | 143 |
| 8.9. | Análisis de las entradas usando Arena | 147 |
| | Ejercicios | 151 |

| | |
|---|------------|
| 9. Generación de números aleatorios | 153 |
| 9.1. Introducción | 153 |
| 9.2. Generadores físicos | 155 |
| 9.3. Generadores aritméticos | 159 |
| 9.4. Generadores congruenciales lineales | 161 |
| 9.5. Otros generadores congruenciales | 166 |
| 9.6. Generadores combinados | 166 |
| 9.7. Generadores de Tausworthe | 168 |
| 9.8. Tests empíricos | 170 |
| 9.9. Tests empíricos de uniformidad | 172 |
| 9.10. Tests empíricos de independencia | 175 |
| Ejercicios | 185 |
| 10. Observaciones de variables aleatorias | 187 |
| 10.1. Introducción | 187 |
| 10.2. Método de la transformación inversa | 188 |
| 10.3. Composición | 191 |
| 10.4. Convolución | 192 |
| 10.5. Método de aceptación/rechazo | 192 |
| 10.6. Observaciones de variables aleatorias continuas | 193 |
| 10.7. Observaciones de variables aleatorias discretas | 195 |
| 10.8. Generación de procesos de llegada | 196 |
| Ejercicios | 198 |
| IV Empleo de los modelos de simulación | 201 |
| 11. Análisis de los resultados de la simulación | 203 |
| 11.1. Introducción | 203 |
| 11.2. Simulaciones con condición de finalización | 208 |
| 11.3. Simulaciones en el estacionario | 209 |
| 11.4. Intervalos para varias magnitudes | 214 |
| 11.5. Comparación entre dos sistemas | 215 |
| 11.6. Comparación entre más de dos sistemas | 220 |
| 11.7. Análisis usando Arena | 221 |
| Ejercicios | 230 |
| 12. Reducción de la varianza | 233 |
| 12.1. Introducción | 233 |
| 12.2. Números pseudo aleatorios comunes | 234 |
| 12.3. Variables aleatorias antitéticas | 240 |
| 13. Diseño de experimentos y optimización | 243 |

| | |
|---|-----|
| 13.1. Introducción | 243 |
| 13.2. Objetivos, respuestas y factores | 245 |
| 13.3. Diseño experimental factorial completo | 248 |
| 13.4. Diseños experimentales factoriales fraccionales | 254 |
| 13.5. Optimización | 256 |
| 13.6. Optimización usando Arena | 257 |
| Ejercicios | 260 |

V Apéndices 263

| | |
|--|------------|
| A. Conceptos de probabilidad y estadística | 265 |
| A.1. ¿Para qué? | 265 |
| A.2. El significado de “probabilidad” | 265 |
| A.3. Probabilidad condicionada | 267 |
| A.4. El teorema fundamental de la probabilidad | 267 |
| A.5. El concepto de variable aleatoria | 268 |
| A.6. Probabilidad acumulada y densidad de probabilidad | 269 |
| A.7. Cuantiles y mediana | 269 |
| A.8. Media | 270 |
| A.9. Varianza y desviación estándar | 270 |
| A.10. Estimadores de la media y de la varianza | 271 |
| A.11. Covarianza y coeficiente de correlación | 271 |
| A.12. Secuencias de variables aleatorias | 273 |
| A.13. Teorema del límite central | 273 |
| A.14. Intervalos de confianza para la media | 274 |
| A.15. Intervalo de confianza para la varianza | 277 |
| A.16. Intervalo de una distribución binomial | 278 |
| A.17. Contraste de hipótesis paramétricas | 280 |
| A.18. Contraste respecto a la media | 281 |
| A.19. Contraste respecto a la varianza | 282 |
| A.20. Contraste respecto a la proporción | 283 |
| A.21. Boxplot (Box & Whiskers) | 283 |
| B. Distribuciones de probabilidad continuas | 285 |
| Uniforme | 285 |
| Exponencial | 286 |
| Gamma | 287 |
| Weibull | 288 |
| Normal | 289 |
| Lognormal | 290 |
| Beta | 291 |

| | |
|--|------------|
| Triangular | 292 |
| Equivalencias entre distribuciones | 293 |
| Relaciones entre distribuciones | 294 |
| C. Distribuciones de probabilidad discretas | 295 |
| Bernoulli | 296 |
| Uniforme discreta | 297 |
| Binomial | 298 |
| Geométrica | 299 |
| Binomial negativa | 300 |
| Poisson | 301 |
| Equivalencias entre distribuciones | 302 |
| Relaciones entre distribuciones | 302 |
| | |
| VI Tablas | 303 |
| | |
| T.1. Puntos críticos de la distribución t y de la distribución normal estándar | 305 |
| | |
| T.2. Puntos críticos de la distribución chi-cuadrado | 306 |
| | |
| T.3. Puntos críticos del test de Kolmogorov-Smirnov | 307 |

Parte I

Introducción al modelado y la simulación

Tema 1

Conceptos básicos del modelado y la simulación

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los conceptos "sistema", "modelo", "simulación" y "marco experimental".
- Describir y comparar los diferentes tipos de modelos.
- Comparar y reconocer los distintos tipos de modelos matemáticos.

1.1. Sistemas y modelos

En el sentido amplio del término, un *modelo* es

"una representación de un sistema desarrollada para un propósito específico".

Puesto que la finalidad de un modelo es ayudarnos a responder preguntas sobre un determinado sistema, el primer paso en la construcción de un modelo es definir cuál es el sistema y cuáles son las preguntas.

En este contexto, se entiende por *sistema*

"cualquier objeto o conjunto de objetos cuyas propiedades se desean estudiar".

Con una definición tan amplia, *cualquier fuente potencial de datos* puede considerarse un sistema. Algunos ejemplos de sistema son:

- Una planta de fabricación con máquinas, personal, dispositivos de transporte y almacén.
- El servicio de emergencias de un hospital, incluyendo al personal, las salas, el equipamiento y el transporte de los pacientes.
- Una red de ordenadores con servidores, clientes, dispositivos de disco y de cinta, impresoras, etc.
- Un supermercado con control de inventario, cajeros y atención al cliente.

- Un parque temático con atracciones, tiendas, restaurantes, trabajadores, clientes y aparcamientos.

Un procedimiento para conocer el comportamiento de los sistemas es la experimentación. De hecho, éste ha sido el método empleado durante siglos para avanzar en el conocimiento: plantear las preguntas adecuadas acerca del comportamiento de los sistemas y responderlas mediante experimentación. Un *experimento* es

“el proceso de extraer datos de un sistema sobre el cual se ha ejercido una acción externa”.

Por ejemplo, el encargado de un supermercado puede ensayar diferentes procedimientos de control del inventario y de distribución del personal, para determinar qué combinaciones demuestran ser más rentables y proporcionan un mejor servicio.

Cuando es posible trabajar directamente con el sistema real, el método experimental presenta indudables ventajas. Sin embargo, para que los resultados del experimento sean válidos, debe garantizarse que no existen variables ocultas “confundidas” con las variables experimentales. Por ejemplo, continuando con el modelo del supermercado, si la afluencia de público durante el periodo en que se experimenta una estrategia de servicio es significativamente mayor que durante el periodo en que se experimenta la otra, no podrá extraerse ninguna conclusión válida sobre el tiempo medio de espera de los clientes en la cola de las cajas.

El *método experimental* está basado en sólidos fundamentos científicos, sin embargo tiene sus limitaciones, ya que en ocasiones es imposible o desaconsejable experimentar con el sistema real. En estos casos, el modelado y la simulación son las técnicas adecuadas para el análisis de sistema, puesto que,

a excepción de la experimentación con el sistema real, la simulación es la única técnica disponible que permite analizar sistemas arbitrarios de forma precisa, bajo diferentes condiciones experimentales.

Existen múltiples razones por las cuales la experimentación con el sistema real puede resultar inviable:

- Quizá la más evidente de ellas es que el sistema aun no exista físicamente. Esta situación se plantea frecuentemente en la fase de diseño de nuevos sistemas, cuando el ingeniero necesita predecir el comportamiento de los mismos antes de que sean construidos.
- Otra posible razón es el elevado coste económico del experimento. Consideremos el caso de un empresario que planea ampliar las instalaciones de su empresa, pero que no está seguro de que la ganancia potencial justifique el coste de la ampliación. Ciertamente, no sería razonable que, para salir de dudas, realizara la ampliación y luego se volviera atrás si ésta demostrara no ser rentable. Una alternativa consiste en simular la operación de la configuración actual de la empresa, simular la operación de la configuración alternativa y comparar sus productividades.
- El experimento puede producir perjuicio o incomodidad. Por ejemplo, experimentar con un nuevo sistema de facturación en un aeropuerto puede producir retrasos y problemas imprevisibles que perjudiquen al viajero.
- En ocasiones el tiempo requerido para la realización del experimento lo hace irrealizable. Casos extremos pueden encontrarse en los estudios geológicos o cosmológicos, de evolución de las especies, sociológicos, etc.
- Algunos experimentos son peligrosos, y por tanto es desaconsejable realizarlos. Por ejemplo, sería inapropiado usar el sistema real para adiestrar a los operarios de una central nuclear acerca de cómo deben reaccionar ante situaciones de emergencia.
- En ocasiones el experimento requiere modificar variables que en el sistema real o bien no están accesibles o no pueden ser modificadas en el rango requerido. Con un modelo matemático adecuado, se pueden ensayar condiciones de operación extremas que son impracticables en el sistema real.

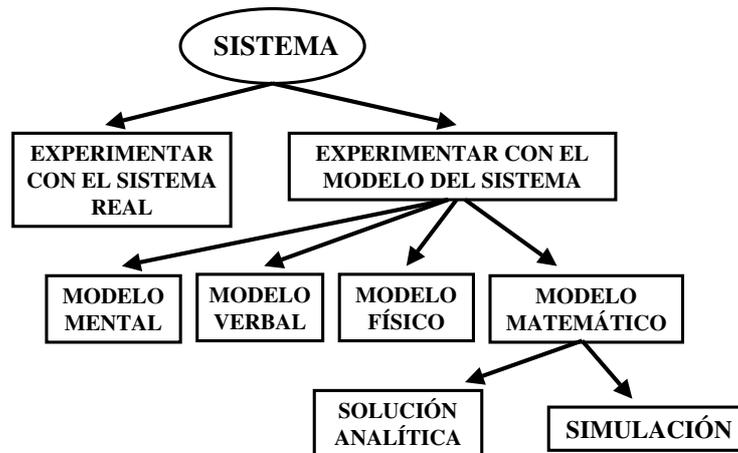


Figura 1.1: Formas de estudiar un sistema.

1.2. Tipos de modelos

Los seres humanos, en nuestra vida cotidiana, empleamos continuamente modelos para comprender y predecir el comportamiento de sistemas (ver la Figura 1.1). Por ejemplo, considerar que alguien es “amable” constituye un modelo del comportamiento de esta persona. Este modelo nos ayuda a responder, por ejemplo, a la pregunta: “¿cómo reaccionará si le pedimos un favor?”. También disponemos de modelos de los sistemas técnicos que están basados en la intuición y en la experiencia. Todos estos se llaman *modelos mentales*.

Por ejemplo, aprender a conducir un coche consiste parcialmente en desarrollar un modelo mental de las propiedades de la conducción del coche. Asimismo, un operario trabajando en determinado proceso industrial sabe cómo el proceso reacciona ante diferentes acciones: el operario, mediante el entrenamiento y la experiencia, ha desarrollado un modelo mental del proceso.

Otro tipo de modelos son los *modelos verbales*, en los cuales el comportamiento del sistema es descrito mediante palabras: si se aprieta el freno, entonces la velocidad del coche se reduce. Los sistemas expertos son ejemplos de modelos verbales formalizados. Es importante diferenciar entre los modelos mentales y los verbales. Por ejemplo, nosotros usamos un modelo mental de la dinámica de la bicicleta cuando la conducimos, sin embargo no es sencillo convertirlo a un modelo verbal.

Además de los modelos mentales y verbales, existe otro tipo de modelos que tratan de imitar al sistema real. Son los *modelos físicos*, como las maquetas a escala que construyen los arquitectos, diseñadores de barcos o aeronaves para comprobar las propiedades estéticas, aerodinámicas, etc.

Finalmente, existe un cuarto tipo de modelos, los *modelos matemáticos*. En ellos, las relaciones entre las cantidades que pueden ser observadas del sistema (distancias, velocidades, flujos, etc.) están descritas mediante relaciones matemáticas. En este sentido, la mayoría de las leyes de la naturaleza son modelos matemáticos.

Por ejemplo, para el sistema “masa puntual”, la Ley de Newton del movimiento describe la relación entre la fuerza y la aceleración. Asimismo, para el sistema “resistencia eléctrica”, la Ley de Ohm describe la relación entre la caída de tensión y el flujo de corriente.

En algunos casos, las relaciones matemáticas que constituyen los modelos son sencillas y pueden resolverse analíticamente. Sin embargo, en la mayoría de los casos, los modelos

no pueden resolverse analíticamente y deben estudiarse, con ayuda del ordenador, aplicando métodos numéricos.

Este experimento numérico realizado sobre el modelo matemático, recibe el nombre de *simulación*.

1.3. Modelos matemáticos

La finalidad de un estudio de simulación (es decir, las preguntas que debe responder) condiciona las hipótesis empleadas en la construcción del modelo, y éstas a su vez determinan qué tipo de modelo resulta más adecuado al estudio. De hecho, un mismo sistema puede ser modelado de múltiples formas, empleando diferentes tipos de modelos, dependiendo de la finalidad perseguida en cada caso.

Existen diferentes clasificaciones de los modelos matemáticos, atendiendo a diferentes criterios. A continuación se describen algunas de las clasificaciones más comúnmente usadas.

Determinista vs Estocástico

Un modelo matemático es *determinista* cuando todas sus variables de entrada son deterministas, es decir, el valor de cada una de ellas es conocido en cada instante.

Un ejemplo de modelo determinista es un servicio al cual los clientes acceden ordenadamente, cada uno a una hora preestablecida (de acuerdo, por ejemplo, con un libro de citas), y en el cual el tiempo de servicio a cada cliente está igualmente preestablecido de antemano. No existe incertidumbre en la hora de inicio o de finalización de cada servicio.

Por el contrario, un modelo es *estocástico* cuando alguna de sus variables de entrada es aleatoria. Las variables del modelo calculadas a partir de variables aleatorias son también aleatorias. Por ello, la evolución de este tipo de sistemas debe estudiarse en términos probabilísticos.

Por ejemplo, considérese el modelo de parking, en el cual las entradas y salidas de coches se producen en instantes de tiempo aleatorios. La aleatoriedad de estas variables se propaga a través de la lógica del modelo, de modo que las variables dependientes de ellas también son aleatorias. Este sería el caso, por ejemplo, del tiempo que transcurre entre que un cliente deja aparcado su vehículo y lo recoge (tiempo de aparcamiento), el número de vehículos que hay aparcados en un determinado instante, etc.

Es importante tener en cuenta que realizar una única réplica de una simulación estocástica es equivalente a realizar un experimento físico aleatorio una única vez.

Por ejemplo, si se realiza una simulación del comportamiento del parking durante 24 horas, es equivalente a observar el funcionamiento del parking real durante 24 horas. Si se repite la observación al día siguiente, seguramente los resultados obtenidos serán diferentes, y lo mismo sucede con la simulación: si se realiza una segunda réplica independiente de la primera, seguramente los resultados serán diferentes.

La consecuencia que debe extraerse de ello es que el diseño y el análisis de los experimentos de simulación estocásticos debe hacerse teniendo en cuenta esta incertidumbre en los resultados, es decir, debe hacerse empleando técnicas estadísticas.

Las dificultades asociadas a la simulación de modelos estocásticos pueden invitarnos en ocasiones a realizar hipótesis adicionales, con el fin de eliminar la incertidumbre en el valor de las variables de entrada. Un ejemplo típico consiste en sustituir cada variable de entrada aleatoria por otra determinista, cuyo valor sea la media de la distribución de probabilidad de aquella. Este modelo determinista, obtenido de eliminar la incertidumbre en el valor de las variables de entrada, proporcionará resultados no aleatorios. Sin embargo, esta simplificación hará que probablemente el modelo ya no sea una representación del sistema válida para el objetivo del estudio.

Estático vs Dinámico

Un *modelo de simulación estático* es una representación de un sistema en un instante de tiempo particular, o bien un modelo que sirve para representar un sistema en el cual el tiempo no juega ningún papel. Ejemplo de simulaciones estáticas son las simulaciones de Monte Carlo.

Por otra parte, un *modelo de simulación dinámico* representa un sistema que evoluciona con el tiempo.

De tiempo continuo vs De tiempo discreto vs Híbrido

Un *modelo de tiempo continuo* está caracterizado por el hecho de que el valor de sus variables de estado puede cambiar infinitas veces (es decir, de manera continua) en un intervalo finito de tiempo. Un ejemplo es el nivel de agua en un depósito.

Por el contrario, en un *modelo de tiempo discreto* los cambios pueden ocurrir únicamente en instantes separados en el tiempo. Sus variables de estado pueden cambiar de valor sólo un número finito de veces por unidad de tiempo.

Pueden definirse modelos con algunas de sus variables de estado de tiempo continuo y las restantes de tiempo discreto. Este tipo de modelos, con parte de tiempo continuo y parte de tiempo discreto, se llama *modelos híbridos*.

Tal como se ha indicado al comienzo de la sección, la decisión de realizar un modelo continuo o discreto depende del objetivo específico del estudio y no del sistema en sí. Un ejemplo de ello lo constituyen los modelos del flujo de tráfico de vehículos. Cuando las características y el movimiento de los vehículos individuales son relevantes puede realizarse un modelo discreto. En caso contrario, puede resultar más sencillo realizar un modelo continuo.

En este punto es conveniente realizar una consideración acerca de los modelos de tiempo continuo y discreto. Al igual que las *variables continuas* (aquellas que pueden tomar cualquier valor intermedio en su rango de variación) son una idealización, también lo son los *modelos de tiempo continuo*. Cualquiera que sea el procedimiento de medida que se emplee para medir el valor de una variable, tendrá un límite de precisión. Este límite marca la imposibilidad de dar una medida continua y supone que, en la práctica, todas las medidas son discretas. Igualmente, los ordenadores trabajan con un número finito de cifras decimales. Sin embargo, para los razonamientos teóricos, conviene considerar ciertas variables como continuas.

Al simular mediante un computador digital un modelo de tiempo continuo, debe discretizarse el eje temporal a fin de evitar el problema de los infinitos cambios en el valor de los estados. Esta discretización constituye una aproximación (con su error asociado) que transforma el modelo de tiempo continuo en un modelo de tiempo discreto. Por ejemplo, si se discretiza el eje temporal del modelo de tiempo continuo

$$\frac{dx}{dt} = f(x, u, t) \quad (1.1)$$

con un intervalo de discretización Δt , se obtiene (empleando el método de Euler explícito) el siguiente modelo de tiempo discreto:

$$\frac{x_{K+1} - x_K}{\Delta t} = f(x_K, u_K, t_K) \longrightarrow x_{K+1} = x_K + \Delta t \cdot f(x_K, u_K, t_K) \quad (1.2)$$

1.4. El marco experimental

Al igual que se distingue entre el sistema real y el experimento, es conveniente distinguir entre la *descripción del modelo* y la *descripción del experimento*. Esto es así tanto desde el punto de vista conceptual como práctico. Sin embargo, en el caso del modelo esta separación

implica cierto riesgo: su empleo en unas condiciones experimentales para las cuales no es válido.

Por supuesto, cuando se trabaja con sistemas reales este riesgo nunca existe, ya que un sistema real es válido para cualquier experimento. Por el contrario, cualquier modelo está fundamentado en un determinado conjunto de hipótesis. Cuando las condiciones experimentales son tales que no se satisfacen las hipótesis del modelo, éste deja de ser válido. Para evitar este problema, la descripción del modelo debe ir acompañada de la documentación de su *marco experimental*. Éste establece el conjunto de experimentos para el cual el modelo es válido.

EJERCICIOS

- 1.1** Describa cuál sería en su opinión la forma más eficaz de estudiar cada uno de los sistemas siguientes, en términos de las posibilidades mostradas en la Figura 1.1.
1. Un ecosistema compuesto por varias especies animales y vegetales, y por recursos (agua, luz, etc.).
 2. Una glorieta en la que convergen varias calles, y que frecuentemente presenta atascos.
 3. Una presa para el suministro de agua y electricidad, que se planea construir en un río.
 4. El servicio de urgencias de un hospital, que se encuentra en funcionamiento.
 5. Un servicio de entrega de pizzas a domicilio.
 6. Una determinada secuencia de pasos en el proceso de fabricación de circuitos integrados, en una fábrica que se encuentra en funcionamiento.
 7. El funcionamiento de un autobús, que conecta el punto de devolución de vehículos, de una compañía de alquiler de coches, con el aeropuerto.
 8. Un circuito eléctrico.
- 1.2** Para cada uno de los sistemas mencionados en el problema anterior, suponga que se ha decidido realizar el estudio mediante simulación. Discuta si la simulación debería ser estática o dinámica, determinista o estocástica, y continua o discreta.
- 1.3** Sugiera al menos dos estudios de simulación en el ámbito de la fabricación/manufactura, y otro en el ámbito de los servicios (por ejemplo, servicios de salud, bomberos, etc.). Justifique el uso de simulación, en oposición al empleo de modelos analíticos. ¿Qué aspectos del análisis mediante simulación son particularmente ventajosos para la aplicación seleccionada?

Tema 2

Pasos en un estudio de simulación

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los pasos de que consta un estudio de simulación.

2.1. Introducción

A grandes rasgos, los pasos que deben seguirse en un estudio de simulación son los siguientes:

1. **Definición del problema.** Consiste en definir claramente por qué se estudia el sistema y qué objetivos se persigue con ello, es decir, cuáles son las preguntas a contestar.
2. **Planificación del proyecto.** Tiene como objetivo estimar qué recursos son necesarios para llevar a cabo el estudio: dinero, personal, hardware de computación, recursos software, etc. Si estos recursos no están disponibles debe replantearse el alcance del estudio.
3. **Definición del sistema.** Debe definirse qué aspectos de la realidad constituyen el sistema bajo estudio.
4. **Formulación conceptual del modelo.** Desarrollo de un modelo preliminar, bien gráficamente (mediante diagramas de bloques) o en pseudo código, en el que se definen los componentes, las variables descriptivas y las interacciones (la lógica) que constituyen el modelo del sistema.
5. **Diseño preliminar del experimento.** Consiste en definir qué acción se va a ejercer sobre el modelo y cómo se va a medir su comportamiento. Se trata, por tanto, de definir qué variables son las entradas y cuáles las salidas, cómo va a modificarse el valor de las entradas y cómo van a recogerse los datos de salida.
6. **Preparación de los datos de entrada.** Debe establecerse el procedimiento a seguir para asignar valores a cada una de las variables de entrada durante la simulación.
7. **Traducción del modelo.** Consiste en describir las partes del modelo y su funcionamiento empleando un lenguaje de simulación. De este modo ya puede ejecutarse la simulación.
8. **Verificación y validación.** Se trata de estudiar si el modelo opera como debiera y si la salida del modelo es creíble y representativa del comportamiento del sistema.

9. **Diseño experimental final.** En este punto se diseña un experimento que proporcione la información necesaria para poder contestar a las preguntas planteadas en el estudio. Típicamente el experimento consiste en cierto número de réplicas de la simulación realizadas en condiciones bien definidas.
10. **Experimentación.** Realización del experimento de simulación diseñado anteriormente.
11. **Análisis e interpretación.** Consiste en inferir conclusiones a partir de los datos obtenidos de la simulación.
12. **Implementación y documentación.** Una vez obtenidos los resultados, analizados y extraídas las conclusiones, éstas deben ser presentadas de la forma más adecuada. El modelo debe ser documentado y preparado de modo que pueda ser reutilizado para posteriores estudios. Finalmente, deben tomarse acciones (si ésta era la finalidad del estudio) a partir de las conclusiones obtenidas.

Respecto a la distribución del esfuerzo, un posible criterio¹ es la regla 40-20-40, consistente en:

- Emplear el 40 % del esfuerzo y del tiempo del proyecto en los pasos 1 a 6 (comprensión del problema, diseño del modelo y del experimento).
- Emplear el 20 % del esfuerzo en el paso 7 (programación del modelo).
- Finalmente, emplear el 40 % restante en los pasos 8 a 12 (verificación, validación, diseño y realización del experimento, análisis de los resultados, implementación y documentación).

A continuación se discute en detalle cada uno de estos pasos.

2.2. Definición del problema y planificación del proyecto

Antes de intentar resolver un problema, es esencial formularlo adecuadamente. Ésta es una afirmación que parece evidente, de sentido común, pero que no siempre se aplica. Para encontrar una solución a un problema, uno debe conocer primero cuál es el problema. En particular, el primer paso en todo estudio de simulación es definir explícitamente cuál es su objetivo: la construcción del modelo se realiza con el fin de ayudar a responder ciertas preguntas acerca del sistema real.

El objetivo de un estudio de simulación puede ser de lo más diverso. Algunos ejemplos son:

- **Evaluación** del comportamiento de determinado diseño del sistema en términos absolutos frente a determinados criterios específicos.
- **Comparación** del comportamiento de diferentes diseños alternativos del sistema, o comparación de diferentes políticas de operación o procedimientos.
- **Predicción** del comportamiento del sistema bajo cierto conjunto de condiciones.
- **Análisis de la sensibilidad**, es decir, determinar entre varios factores cuál es el que más afecta al comportamiento del sistema.
- **Optimización**, esto es, cálculo de los valores de determinados parámetros del modelo que producen que el sistema tenga la mejor respuesta (atendiendo a un determinado criterio).
- **Estudio de las relaciones funcionales** entre determinadas variables del modelo. Por ejemplo, la determinación de la relación funcional entre una de las respuestas del modelo y determinadas variables de entrada.
- **Localización y análisis de los cuellos de botella.** Consiste en descubrir la localización de los cuellos de botella del sistema y evaluar propuestas para descongestionar alguno de ellos.

Estos objetivos pretenden responder preguntas del tipo:

- ¿Funciona el sistema del modo que queremos que lo haga (evaluación y predicción)?

¹(Pedgen, Shannon & Sadowsky 1995)

- ¿Por qué no funciona el sistema y qué podemos hacer para que funcione (determinación de los cuellos de botella y optimización)?
- ¿Cuál es la mejor alternativa (comparación y análisis de sensibilidad)?

El propósito del estudio de simulación tiene implicaciones decisivas en el diseño del modelo y del experimento. Por ejemplo, si el propósito es evaluar el comportamiento de un sistema en términos absolutos, deberá existir un alto grado de correspondencia entre el comportamiento del modelo y del sistema. Por el contrario, si el propósito del estudio es comparar varios diseños, el modelo puede ser válido en un sentido relativo incluso cuando sus respuestas en un sentido absoluto difieran considerablemente de las del sistema real. En consecuencia, el proceso completo de diseño del modelo, validación, diseño de los experimentos y análisis de los resultados debe estar íntimamente relacionado con el propósito del estudio.

No obstante, el objetivo del estudio debe definirse con más detalle que mediante una simple pregunta. También es necesario conocer todos los aspectos relevantes que rodean al proyecto de simulación. Frecuentemente el estudio es realizado por un grupo de ingenieros que deben recomendar una solución al problema, pero que no son quienes deben tomar la decisión final. En estos casos, los resultados del estudio deben servir simultáneamente a dos propósitos: ayudar a formular la recomendación y ayudar a “vender” la recomendación. Con este fin, deben considerarse otros factores condicionantes: cuáles son los intereses de la persona que debe tomar la decisión final, qué aspectos de la situación están bajo su control y cuáles no, en qué va a consistir el proceso de la toma de decisión, qué otras personas participan en la toma de la decisión (en especial aquellas que se oponen a que se realicen cambios en el sistema) y cuáles son sus objetivos más relevantes, etc.

Como parte de la planificación deben considerarse todos aquellos factores que son críticos para el éxito del proyecto, tales como:

- ¿Están claramente definidos los objetivos del estudio? ¿Por qué se realiza el estudio? ¿Qué se pretende obtener del estudio?
- ¿Se dispone de los recursos adecuados y suficientes para poder realizar el estudio: tiempo, personal, software, ordenadores, etc.?
- ¿Dispone el estudio del soporte necesario por parte de la dirección? Las partes involucradas se implicarán en mayor medida si son conscientes del respaldo de la dirección. ¿Se les ha informado adecuadamente? ¿Son conscientes de que ese respaldo existe?
- ¿Se han definido las tareas de que consta el proyecto, la vinculación entre ellas, quién debe realizarlas y en qué fechas?
- ¿Posee cada uno de los miembros del proyecto la formación necesaria para realizar con éxito su cometido? ¿Está suficientemente capacitado el coordinador del proyecto?
- ¿Se han identificado los usuarios potenciales o clientes de los resultados del estudio? ¿Han sido todos ellos consultados? ¿Se ha establecido el procedimiento para comunicarles regularmente el estado en que se encuentra el proyecto y los progresos realizados?
- ¿Se han establecido los adecuados canales de comunicación entre los miembros del equipo, la dirección y los clientes?

2.3. Definición del sistema y formulación del modelo

La esencia del arte del modelado son la *abstracción* y la *simplificación*. Se trata de identificar el pequeño conjunto de características o propiedades del sistema suficientes para servir los objetivos específicos del estudio. A grandes rasgos, la metodología para la construcción del modelo podría ser la siguiente:

- Escoger las variables de salida, lo cual resulta relativamente sencillo una vez definido el objetivo del estudio.
- La construcción del modelo se realiza identificando qué componentes del sistema afectan a estas variables de salida y decidiendo, para cada uno de ellos, si debe ser incluido en

el modelo o si debe ser considerado parte del entorno del modelo. En este último caso, el componente es representado mediante entradas al modelo.

- Una vez se ha decidido qué componentes constituyen de modelo y qué variables describen el comportamiento de cada componente, deben determinarse las relaciones funcionales entre los componentes, es decir, la lógica del modelo. Normalmente se usan gráficos o pseudo código para describir el sistema como un diagrama de flujo lógico.

Para diseñar el modelo de simulación conviene disponer de la máxima información posible acerca del sistema. Existen diversas fuentes de información, por ejemplo:

- *La documentación de las especificaciones de diseño del sistema.* Aunque posiblemente esta información no haya sido actualizada desde hace tiempo y esté incompleta, suele ser un punto de partida razonable para comprender el sistema.
- *Entrevistas con operarios e ingenieros* que conozcan el sistema o con sus diseñadores si éste todavía no ha sido construido. En efecto, durante el diseño del modelo puede invitarse a especialistas en el sistema a que examinen datos de salida, procedentes unos del sistema real y otros del modelo, y que intenten, sin conocer la respuesta, identificar (razonadamente) cuáles provienen del modelo y cuáles del sistema real. Esta información puede emplearse para mejorar el modelo.
- *Observaciones y medidas* realizadas “de primera mano” sobre el sistema. Sin duda ésta es la fuente de información más fiable: una vez leída la documentación y discutida la operación del sistema con operarios e ingenieros, es recomendable observar el sistema en operación. De hecho, en los primeros años de la investigación de operaciones se consideraba un procedimiento habitual que los ingenieros a cargo del modelado participaran durante cierto tiempo en la operación del sistema bajo estudio a fin de conocerlo mejor.

La tarea de modelado implica la búsqueda de un punto de equilibrio: el modelo debe representar los aspectos del sistema real con el grado de detalle requerido, pero de modo que sea lo más sencillo posible. No debe hacerse un modelo demasiado detallado, que incluya componentes que contribuyan poco o nada a la comprensión del problema. Debe tenerse presente que el objetivo del modelo es responder a las preguntas planteadas y no imitar de manera precisa el comportamiento del sistema real. Además, aumentando la complejidad del modelo, no necesariamente se consigue aumentar su realismo.

Ejemplo 2.3.1. *En (Law & Kelton 2000) se describe el siguiente caso real. La dirección de una fábrica de comida para perros encargó a una consultoría la realización de un modelo de su línea de producción, que producía 1 millón de latas de comida al día, a velocidad constante. Al realizar el modelo, se representó cada lata de comida como una entidad independiente, con lo cual era extremadamente costoso ejecutar el modelo, y por ello, no fue muy útil. Unos años más tarde, el modelo fue reescrito, tratando el proceso de fabricación como un “flujo continuo”. El nuevo modelo produjo resultados precisos, y se ejecutaba en una pequeña fracción del tiempo necesario para ejecutar el modelo original.* □

Ejemplo 2.3.2. *En (Law & Kelton 2000) se describe otro caso real. En 1985 se realizó un modelo de una fábrica de grandes dimensiones (media 1.5 millas de largo), con un coste de \$250,000. Sin embargo, el modelo era tan detallado que jamás pudo llegarse a ejecutar, dado los inmensos recursos de memoria que requería.* □

Una buena práctica consiste en realizar el modelo de manera iterativa: comenzar con un modelo muy simple, cuya complejidad puede posteriormente ir aumentándose fácilmente. Para ello, el modelo debe realizarse de manera *modular* y *jerárquica*, dividiendo el sistema en submodelos y modelando todos ellos con un nivel semejante de complejidad. Este modelo inicial puede construirse muy rápido y puede servir de punto de discusión sobre posteriores refinamientos en el modelado (se entiende por “refinamiento” del modelo el aumento en su nivel de detalle). Añadiendo progresivamente los refinamientos al modelo, y comparando los resultados obtenidos con los del modelo más sencillo, puede estimarse el impacto de cada conjunto de refinamientos sobre la respuesta del modelo. En determinado punto de este proceso de aumento gradual de la complejidad del modelo, los refinamientos añadidos tienen

un efecto pequeño, es decir, influyen despreciablemente en las conclusiones del estudio, con lo cual se concluirá que no es preciso incorporarlos.

Existen varias técnicas que se pueden aplicar simultáneamente para obtener un modelo simplificado de un sistema. Algunas de ellas son:

- La **omisión** de ciertos detalles del modelo. La idea que hay detrás de esta aproximación es que no todos los factores son igualmente importantes para determinar el comportamiento del sistema. Se trata de determinar qué factores son críticos (efectos de primer orden) y cuáles no (efectos de segundo orden).

Omitir los detalles supone introducir deliberadamente “imperfecciones”, y por tanto podría considerarse que es una mala práctica. Sin embargo, nada más lejos de la realidad: solamente omitiendo los factores de segundo orden puede construirse un modelo útil. El aumento en el nivel de detalle debe hacerse atendiendo al impacto que la complejidad añadida tiene sobre los resultados del estudio, es decir, debe emplearse el esfuerzo de modelado en la representación de aquellos aspectos del sistema que más influencia tienen sobre los resultados del estudio.

- **Agregar** determinados procesos del sistema en un proceso equivalente. Ejemplos típicos de agregación son:
 - *Agregación temporal*: se trata un intervalo de tiempo como una unidad. Por ejemplo, se consideran que todos los eventos ocurridos en un día han ocurrido simultáneamente en determinado instante (por ejemplo, a las 12 de la mañana).
 - *Agregación de recursos*: se tratan varios recursos como una unidad. Por ejemplo, si un operario de una cadena de montaje realiza cierto número de trabajos consecutivos, puede modelarse la operación completa como un proceso simple en lugar de modelar cada una de las tareas individualmente.

El Teorema del Límite Central a menudo ayuda a determinar la distribución de probabilidad resultante del proceso agrupado. Si la variable aleatoria del proceso agrupado (por ejemplo, el tiempo en realizar la tarea) es suma de las variables aleatorias de los subprocesos y si éstas son independientes entre sí, entonces la variable del proceso agrupado está aproximadamente distribuida de forma normal cuando el número de subprocesos es grande (sean cuales sean las distribuciones de probabilidad de las variables de los subprocesos).

- El tercer tipo de simplificación consiste en **sustituir** un proceso complejo por un proceso diferente, más sencillo, que se comporte de manera equivalente al proceso original. Un ejemplo es la representación de un proceso complejo, cuyo comportamiento no se comprende o no se desea modelar, por medio de determinada distribución de probabilidad. De este modo, se sustituye un componente del sistema por entradas al sistema. Por ejemplo, para modelar los pedidos realizados por determinado cliente puede, o bien modelarse toda la estructura de negocio del cliente, o bien considerar que sus demandas están distribuidas aleatoriamente de acuerdo a determinada distribución de probabilidad.

Además de la aplicación de las técnicas anteriores, comúnmente se realizan las siguientes aproximaciones:

- La definición de los “límites” del modelo y de su interacción con el entorno implica la realización de aproximaciones, entre ellas la decisión de qué factores externos influyen sobre el funcionamiento del modelo (entradas al modelo) y cuáles son ignorados.
- Los fenómenos de naturaleza aleatoria son modelados mediante funciones de probabilidad (frecuentemente distribuciones teóricas) que los representan sólo de manera aproximada.
- Al modelar la interacción entre los fenómenos aleatorios, en ocasiones se considera (de forma aproximada) que dos variables aleatorias son estadísticamente independientes con el fin de simplificar el modelo.
- En ocasiones se considera que la distribución de probabilidad de algunos de los parámetros del modelo no varía con el tiempo. Esta aproximación es razonable si la velocidad de cambio es muy pequeña comparada con el periodo de tiempo de interés.

El impacto de las aproximaciones de modelado debe estimarse considerando el propósito específico del estudio de simulación:

- Si el propósito es estimar de manera precisa el comportamiento del sistema, deben evaluarse las simplificaciones considerando su impacto sobre la medida del comportamiento.
- Por otro lado, si el propósito del estudio es comparar diferentes sistemas o procedimientos, debe evaluarse si la simplificación afecta a todos los modelos aproximadamente de la misma manera (los errores son aproximadamente iguales en magnitud).

Por este motivo, estimar el comportamiento absoluto de un sistema normalmente requiere modelos considerablemente más detallados que comparar varios sistemas. Una consecuencia práctica de este hecho es que los modelos matemáticos suelen ser más precisos en la comparación de sistemas alternativos que en la obtención de respuestas absolutas.

2.4. Diseño de los experimentos

El diseño de los experimentos se realiza en dos etapas diferentes del estudio de simulación. En primer lugar, antes de que el diseño del modelo haya finalizado. Tan pronto como sea posible deben seleccionarse qué medidas del comportamiento del sistema van a usarse en el estudio, qué factores van a variarse y qué niveles de cada uno de estos factores van a investigarse. Una vez que el modelo ha sido desarrollado, verificado y validado, se revisa el diseño experimental teniendo en cuenta el conocimiento que se ha ganado durante el proceso de diseño, construcción, verificación y validación del modelo.

El experimento debe diseñarse cuidadosamente, atendiendo a las siguientes dos consideraciones:

- El diseño del experimento determina el tipo de análisis estadístico que debe aplicarse a los resultados.
- Los experimentos de simulación deben diseñarse de modo que permitan obtener la máxima información acerca del comportamiento del sistema al menor costo posible (tiempo y trabajo del experimentador, y tiempo de computación).

2.5. Datos de entrada

Determinar qué datos emplear como entrada al modelo durante la simulación es una tarea muy delicada, de la que depende en gran medida el éxito del estudio. De hecho, frecuentemente la recogida y el análisis de los datos del sistema, y el modelado de las entradas a partir de estos datos, consume más tiempo que el propio diseño y programación del modelo.

El diseño de un modelo estocástico de simulación siempre implica decidir si determinados aspectos del sistema son representados de forma estocástica o determinista. Si se opta por una descripción probabilística de determinado aspecto, puede suceder que:

- Puedan obtenerse datos experimentales del sistema. En este caso, debe decidirse si durante la simulación se muestrea directamente de los datos experimentales, o bien si se muestrea de una distribución de probabilidad ajustada a partir de éstos.
- No puedan obtenerse datos experimentales, bien porque el sistema aun no ha sido construido o porque no es posible recogerlos. En este caso, puede uno basarse en consideraciones teóricas y en estimaciones de operarios, ingenieros, diseñadores, etc. familiarizados con el funcionamiento del sistema.

2.6. Traducción del modelo

Durante la década de los 60 las simulaciones se realizaban usando *lenguajes de programación procedurales*, de propósito general, como es el caso de FORTRAN. Como puede suponerse, la escritura de estos programas resultaba una tarea larga, costosa, tediosa y propensa a errores, ya que el programador debía tener en cuenta hasta el más mínimo detalle de la simulación del modelo. Esta metodología de modelado se denomina *modelado orientado a los eventos*.

Con el fin de poder reutilizar parte del código, se desarrollaron bibliotecas de funciones, que realizaban algunas de las tareas rutinariamente requeridas en una simulación y que podían ser invocadas desde el programa de simulación. Sin embargo, los modelos debían ser programados prácticamente desde el principio cada vez que se introducían en ellos modificaciones importantes. En resumen, la simulación era una técnica muy costosa y especializada, sólo al alcance de empresas capaces de realizar grandes inversiones económicas.

Ante la gran dificultad (algunos autores hablan de imposibilidad) de realizar estudios complejos de simulación basados en ésta metodología, se desarrolló una nueva metodología de modelado: el modelado orientado a los procesos. Esta metodología permite describir el modelo de manera más “natural”, más próxima al razonamiento humano. En la década de los años 70 aparecieron los *lenguajes de simulación* de propósito general para modelos de tiempo discreto, que posibilitaron la práctica de esta nueva metodología. Estos entornos de simulación traducían automáticamente la descripción orientada al proceso del modelo (realizada por el usuario), a la descripción orientada a los eventos, expresada en un determinado lenguaje de programación, cuya ejecución constituye la realización del experimento de simulación. Algunos de estos lenguajes de simulación son todavía usados en la actualidad: GPSS, SIMSCRIPT, SLAM, SIMAN, etc..

Durante la década de los 80 los lenguajes de simulación evolucionaron. Aparecieron nuevas versiones que recogían los dos conceptos fundamentales siguientes:

- La separación en actividades distintas de las etapas (funcionalmente distintas) de modelado, especificación del experimento y análisis de las salidas.
- El aprovechamiento de todas las capacidades disponibles en el momento para el manejo de bases de datos, capacidades gráficas y de verificación del programa.

Los avances experimentados en todos los ámbitos de la computación durante la primera mitad de la década de los 90, posibilitaron el desarrollo de los *entornos de modelado*: una capa software construida sobre el lenguaje de simulación con el fin de facilitarle al usuario la descripción del modelo. El manejo de cualquier lenguaje de simulación requiere un esfuerzo previo por parte del usuario en el aprendizaje de sus particularidades sintácticas. Con el fin de facilitar la descripción del modelo, los entornos de modelado proporcionan interfaces de usuario muy intuitivas con menús, diálogos, etc. El usuario construye el modelo seleccionando los componentes (pinchando y arrastrando el icono) y conectándolos gráficamente. La animación y otras capacidades gráficas permiten visualizar la evolución del sistema durante la simulación. El interfaz gráfico de usuario del entorno de modelado también permite acceder a los niveles inferiores en la descripción del modelo: a la descripción de partes del modelo usando el lenguaje de simulación e incluso al lenguaje de programación.

Un ejemplo es el entorno de modelado Arena, que soporta el uso combinado de diferentes niveles de descripción: elementos de alto nivel parametrizables por el usuario y elementos de bajo nivel que el usuario puede definir usando el lenguaje de simulación SIMAN o los lenguajes de programación Visual Basic o C/C++.

2.7. Verificación y validación del modelo

Los procesos de verificación y la validación son conceptualmente distintos:

- La finalidad de la *verificación* es comprobar que no se han cometido errores al traducir el modelo, bien usando un entorno de modelado o mediante un lenguaje de simulación o de programación.
- La *validación* consiste en comprobar que el modelo supone una aproximación adecuada de la realidad para los objetivos particulares del estudio de simulación.

Sin embargo, cuando los resultados de la simulación nos parecen “extraños” o erróneos, no siempre está claro si es debido a que nos hemos equivocado al traducir el modelo o a que las hipótesis de modelado no son las adecuadas.

Verificación

Entre otros, pueden usarse los siguientes procedimientos para verificar el modelo:

- **Verificación manual de la lógica.** Consiste en ejecutar la simulación durante un periodo de tiempo corto y comprobar manualmente los resultados obtenidos.
- **Comprobación submodelo a submodelo.** Se trata de verificar individualmente que cada submodelo produce los resultados esperados para todos los posibles tipos de entradas.
- **Comprobación con soluciones conocidas.** Consiste en ajustar el modelo de modo que represente un sistema de solución conocida y comparar ésta con los resultados de la simulación.
- **Test de sensibilidad.** Puede modificarse el valor de un parámetro, dejando los demás fijos, con el fin de medir la sensibilidad del modelo respecto a ese parámetro. La comparación de la sensibilidad observada en las simulaciones, con la que sería de esperar en el sistema real, puede proporcionar pistas útiles.

Validación

Puede considerarse que la validación del modelo tiene tres vertientes diferentes. Consiste en determinar:

- Si el modelo representa adecuadamente al sistema real (comprobación de la estructura del modelo).
- Si los datos generados de la simulación del modelo reproducen de forma adecuada el comportamiento del sistema real (comprobación del comportamiento del modelo).
- Si el usuario del modelo tiene confianza en los resultados obtenidos de las simulaciones (comprobación de la confianza del usuario en el modelo). Involucrar al usuario final en todas las fases del diseño y la construcción del modelo generalmente hace que este aspecto de la validación del modelo sea mucho más sencillo.

Ejemplo 2.7.1. *En (Law & Kelton 2000) se describe el siguiente caso real. Un analista militar estadounidense trabajó durante 6 meses en un modelo, sin interactuar con el general que lo había solicitado. En la reunión mantenida en el Pentágono para la presentación de los resultados del estudio, cuando tan sólo habían transcurrido 5 minutos, el general afirmó: “Este no es el problema en el que yo estoy interesado”.*

Puesto que el modelo se construye para un propósito específico, la validez sólo puede ser evaluada con relación a este propósito. La validación del modelo es un proceso continuo durante su diseño, desarrollo y uso. Existen diferentes grados de validación: la confianza en el modelo va acumulándose según el modelo va superando pruebas y se van encontrando más puntos de coincidencia entre el comportamiento del modelo y el del sistema real. La verificación y la validación de un modelo son procesos que realmente nunca finalizan.

En todo este proceso de validación, no debe perderse de vista que el objetivo del ingeniero dedicado al modelado es la realización de modelos útiles, en un tiempo razonable y con un coste razonable. Por este motivo, más que preguntarse en qué medida se ajusta el comportamiento simulado al comportamiento real del sistema, es más adecuado preguntarse en qué medida las diferencias entre el modelo y el sistema son lo suficientemente significativas como para afectar a las conclusiones derivadas del uso del modelo.

2.8. Experimentación y análisis de los resultados

La planificación del experimento y el análisis de los resultados requieren la aplicación de conocimientos estadísticos. El método experimental y el análisis estadístico a aplicar dependen del tipo de simulación: con terminación o sin terminación. El tipo de simulación (con o sin terminación) depende fundamentalmente del objetivo del estudio, más que de la naturaleza en sí del sistema. A continuación se definen estos conceptos:

- La simulación se dice que tiene *terminación* cuando existe un evento que señala de forma natural el final de la misma. La finalidad es estudiar el comportamiento del sistema cuando éste evoluciona, partiendo de unas determinadas condiciones iniciales, hasta que se satisface determinada condición de finalización.
- Por el contrario, se dice que la simulación es *sin terminación* cuando no existe un evento que de manera natural marque el final de la misma. La finalidad de este tipo de estudios es analizar el comportamiento del sistema una vez éste ha adquirido un régimen de funcionamiento independiente de las condiciones iniciales de la simulación, es decir, el estado estacionario.

Puesto que la salida de un modelo de simulación es una muestra de datos de su comportamiento, pueden aplicarse los conceptos de inferencia estadística de muestras. Sin embargo, la mayoría de los tests estadísticos requieren que los datos de la muestra sean independientes entre sí, es decir, no estén correlacionados. Frecuentemente los datos obtenidos de la simulación de los modelos no cumplen esta condición, con lo cual deben aplicarse ciertos procedimientos antes de poder extraer conclusiones válidas. Respecto al tamaño de la muestra (duración de la simulación o número de réplicas), éste debe ser lo suficientemente grande como para proporcionar el nivel de confianza necesario en la estimación de la medida del comportamiento.

2.9. Documentación e implementación

Llegado este punto, ya se han completado todos los pasos del diseño y desarrollo del modelo, ejecución de la simulación, así como del análisis de los resultados. Los elementos finales del estudio de simulación son la implementación y la documentación. No se puede considerar que un proyecto de simulación se ha completado con éxito hasta que los resultados obtenidos no han sido entendidos, aceptados y usados. Las conclusiones del estudio deben ser documentadas de manera clara, concisa y convincente, debiéndose abordar todos aquellos aspectos que el cliente considera importantes. De otra forma, difícilmente serán aplicadas, con lo cual el proyecto será un fracaso. La presentación de los resultados es una parte del estudio tan crítica como las demás, y merece el mismo cuidado en su planificación y diseño.

Además de los resultados del estudio, debe documentarse el desarrollo y operación del modelo. Esto puede contribuir a aumentar considerablemente su vida útil, así como aumentar la probabilidad de que recomendaciones basadas en él sean aceptadas. Una buena documentación facilita las modificaciones y asegura que el modelo puede ser usado o algunas de sus partes reutilizadas, aun si no están presentes sus desarrolladores.

EJERCICIOS

2.1 Plantee un posible estudio de simulación del sistema siguiente: una gasolinera, con varios surtidores, atendida por varios empleados. En particular, responda a las cuestiones siguientes:

- ¿Cuáles son las preguntas a responder?
- ¿Qué recursos prevé que necesitaría para llevar a cabo el estudio? Sugiera un método adecuado de recogida de datos experimentales, justifíquelo y describa cómo lo ejecutaría.

- ¿Qué medidas del comportamiento del sistema son de interés?
 - ¿Qué aspectos de la realidad constituyen el sistema bajo estudio?
 - ¿Cuáles son las variables de entrada del modelo? ¿De qué tipo es cada una: aleatoria o determinista?
 - Describa las hipótesis de modelado.
 - Realice un diseño preliminar del experimento.
- 2.2** Plantee un posible estudio de simulación del sistema siguiente: la intersección de varias calles, que se encuentra regulada por semáforos. Conteste a las cuestiones planteadas en el Problema 2.1.
- 2.3** Plantee un posible estudio de simulación del sistema siguiente: un servicio de ambulancias de un hospital. Conteste a las cuestiones planteadas en el Problema 2.1.

Parte II

Modelado y simulación

Tema 3

Método de Monte Carlo

NOTA IMPORTANTE: Este tema NO SE EXIGIRÁ EN EL EXAMEN durante el presente curso académico, si bien se recomienda su lectura. Por este motivo, no se plantean objetivos docentes del tema, ni tampoco se proponen problemas.

3.1. Introducción

Bajo el nombre genérico de “simulaciones de Monte Carlo” suelen englobarse todas las simulaciones que emplean números aleatorios para resolver problemas estáticos, es decir, problemas en los cuales el tiempo no juega un papel relevante.

El término “Monte Carlo” proviene del nombre en clave de un trabajo secreto en el que von Neumann y Ulam emplearon esta técnica matemática (que ya era conocida anteriormente). Este trabajo se realizó en Los Alamos, durante el conocido proyecto para la fabricación de la bomba atómica.

En esta técnica se generan artificialmente datos mediante el uso de un generador de números aleatorios y la función de probabilidad de interés. El generador de números aleatorios se emplea para generar variables aleatorias distribuidas uniformemente en el intervalo $[0, 1]$. La función de probabilidad a muestrear puede ser una distribución teórica (normal, exponencial, etc.) o puede ser empírica, es decir, puede estar basada en observaciones del sistema real.

La técnica de Monte Carlo se emplea no sólo para simular sistemas estocásticos, sino también para resolver problemas completamente deterministas (típicamente aquellos que no pueden ser resueltos analíticamente).

A continuación se discuten algunos ejemplos de aplicación de la técnica:

- La estimación del valor del número π , mediante la simulación del experimento de la aguja de Buffon.
- La estimación de la probabilidad que hay de ganar un juego “solitario” de cartas.
- El cálculo de la una integral definida del tipo $I = \int_a^b g(x) \cdot dx$, donde $g(x)$ es una función real monovaluada. Es conveniente aclarar que, en aquellos casos en que el integrando $g(x)$ tiene “buen comportamiento” (es decir, ni la función ni sus derivadas presentan discontinuidades), suele ser preferible emplear otros métodos numéricos de integración a emplear el método de Monte Carlo.



Figura 3.1: Retrato de G. L. Leclerc, Conde de Buffon, en un sello francés.

3.2. Estimación del valor del número π

Alrededor de 1733, George Louis Leclerc (su retrato se muestra en la Figura 3.1), describió un experimento para estimar el valor del número π . Si se lanza una aguja de longitud l sobre una mesa pintada con líneas paralelas, con un espaciado entre ellas d (donde $d \geq l$), entonces la probabilidad de que la aguja corte una línea es:

$$p = \frac{2 \cdot l}{\pi \cdot d} \quad (3.1)$$

Puede, por tanto, plantearse el siguiente experimento para estimar el número¹ π :

1. Conseguir una aguja y pintar las líneas sobre la mesa.
2. Decidir cuántas veces va a lanzarse la aguja. Sea N el número de veces.
3. Poner un contador a cero.
4. Repetir N veces las dos acciones siguientes:
 - a) Lanzar la aguja "aleatoriamente" sobre la mesa.
 - b) Si la aguja cruza alguna de las líneas, incrementar el contador en uno. Si la aguja no cruza ninguna línea, el contador no varía.
5. Calcular la proporción de veces que la aguja ha cruzado alguna línea:

$$\hat{p} = \frac{\text{valor final del contador}}{N} \quad (3.2)$$

6. Si $\hat{p} > 0$, entonces estimar el número π de la forma siguiente:

$$\hat{\pi} = \frac{2 \cdot l}{\hat{p} \cdot d} \quad (3.3)$$

Como vemos, el experimento de la aguja de Buffon es un ejemplo de método de Monte Carlo, ya que se están empleando números aleatorios para estimar el valor de cierta propiedad.

Para simular mediante el ordenador este experimento, debemos ser capaces de situar la posición de la aguja aleatoriamente entre las líneas paralelas. Cuando se realiza el experimento físicamente, esto no constituye un problema, ya que simplemente lanzamos la aguja, vemos donde ha caído y anotamos si corta alguna línea o no.

¹Este experimento se conoce como *el problema de la aguja de Buffon*, ya que G. L. Leclerc fue Conde de Buffon.

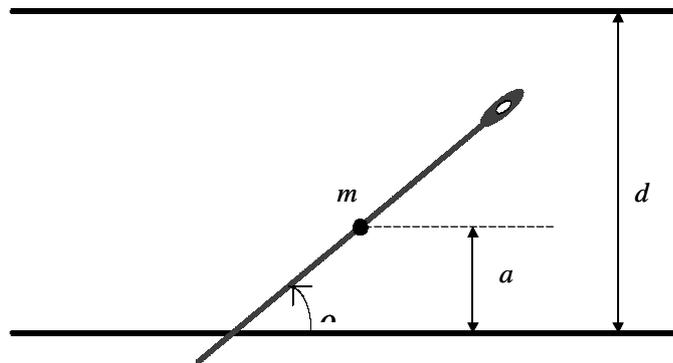


Figura 3.2: Experimento de la aguja de Buffon.

Para determinar la posición de la aguja de manera unívoca, basta con conocer dos datos: la posición en la mesa de algún punto determinado de la aguja, por ejemplo de su centro, m , y el ángulo que forma la aguja respecto a las líneas paralelas, θ .

A partir de las siguientes dos consideraciones, puede obtenerse un procedimiento para determinar si la aguja corta alguna línea:

- Una vez fijada la posición del punto medio de la aguja, m , debe determinarse cuál de las dos líneas es la más cercana a este punto medio. Sea a la distancia entre el punto medio de la aguja y la línea más cercana. Esta distancia a es una variable aleatoria distribuida uniformemente en el intervalo $[0, \frac{d}{2}]$.
- Debido a la simetría, el ángulo que forma la aguja respecto a las líneas paralelas, θ , es una variable aleatoria distribuida uniformemente en el intervalo $[0, \pi]$.

Una vez obtenidos los valores de a y θ , puede seguirse el siguiente procedimiento para decidir si la aguja toca o interseca la línea más próxima al centro de la aguja (ver Figura 3.2):

- Si $a \leq \frac{l}{2} \cdot \sin(\theta)$, entonces la aguja toca o corta la línea.
- Si $a > \frac{l}{2} \cdot \sin(\theta)$, entonces la aguja ni toca ni corta la línea.

Las muestras aleatorias a y θ pueden obtenerse a partir de números seudo aleatorios, u_1 y u_2 , de la forma siguiente:

$$a = \frac{d}{2} \cdot u_1 \quad (3.4)$$

$$\theta = \pi \cdot u_2 \quad (3.5)$$

donde u_1 y u_2 son observaciones de una variable aleatoria uniformemente distribuida en el intervalo $[0, 1]$.

El flujo del programa para simular el problema de la aguja de Buffon es el siguiente:

Paso 1: Inicializar $N_{\text{corta}} = n = 0$, contadores del número de veces que la aguja corta o toca la línea y del número de lanzamientos respectivamente.

Paso 2: Seleccionar l y d . Seleccionar N_{total} , que es el número total de lanzamientos de la aguja.

Paso 3: Generar dos números seudo aleatorios, u_1 y u_2 .

Paso 4: Calcular a y θ , empleando las Ecuaciones (3.4) y (3.5).

Paso 5: Si $a > \frac{l}{2} \cdot \sin(\theta)$, entonces saltar al Paso 7.

Paso 6: Incrementar N_{corta} .

Paso 7: Incrementar n .

Paso 8: Si $n > N_{\text{total}}$, entonces ir al Paso 9. En caso contrario, ir al Paso 3.

Paso 9: Calcular e imprimir el resultado:

$$\hat{\pi} = \frac{2 \cdot l}{\frac{N_{\text{corta}}}{N_{\text{total}}} \cdot d} \quad (3.6)$$

Este algoritmo puede programarse usando cualquier lenguaje de programación. Por ejemplo, ejecutando el programa con las siguientes entradas: $l = 10$, $d = 20$ y $N_{\text{total}} = 3000$, y empleando una determinada secuencia de números aleatorios, se obtiene la siguiente estimación de la probabilidad de que la aguja toque o corte una línea :

$$\hat{p} = \frac{N_{\text{corta}}}{N_{\text{total}}} = 0.3133 \quad (3.7)$$

Para determinar la precisión de esta estimación, debe construirse un intervalo de confianza. Cada lanzamiento de la aguja es un experimento de Bernoulli. Puede considerarse que el número de aciertos es N_{corta} y el número de intentos es N_{total} . Así pues, $\hat{p} = \frac{N_{\text{corta}}}{N_{\text{total}}}$ es un estimador del parámetro binomial p : la probabilidad de éxito. En el Apéndice A se indica cómo construir el intervalo de confianza para la proporción de una distribución binomial. Se obtiene el siguiente intervalo de confianza del 95 %:

$$\text{Prob} \left\{ \hat{p} - z_{0.975} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N_{\text{total}}}} \leq p \leq \hat{p} + z_{0.975} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{N_{\text{total}}}} \right\} = 0.95 \quad (3.8)$$

Sustituyendo los valores obtenidos en la simulación, y teniendo en cuenta que $z_{0.975} = 1.96$ (ver la última fila de la Tabla T.1, al final del libro), se obtiene:

$$\text{Prob} \{0.2967 \leq p \leq 0.3299\} = 0.95 \quad (3.9)$$

El intervalo de confianza para π se calcula teniendo en cuenta que:

$$\hat{\pi} = \frac{2 \cdot l}{\hat{p} \cdot d} = \frac{2 \cdot 10}{\hat{p} \cdot 20} = \frac{1}{\hat{p}} \quad (3.10)$$

Se obtiene el siguiente intervalo de confianza del 95 %:

$$\text{Prob} \{3.031 \leq \pi \leq 3.3704\} = 0.95 \quad (3.11)$$

Este experimento es muy simple, pero hay algunos aspectos de él que son comunes a la mayoría de las simulaciones:

- El propósito es estimar algo (en este caso, el número π) cuyo valor es difícil calcular exactamente (al menos así era en el año 1733).
- El valor estimado que se obtiene del estudio de simulación no es exactamente correcto: existe un cierto error asociado a la estimación y resulta conveniente obtener una estimación de la magnitud de este error.
- Parece intuitivo que cuantos más lanzamientos de la aguja se hagan, menor será el error cometido y, por tanto, mejor será la estimación del número π .
- De hecho, en lugar de decidir de antemano el número total de lanzamientos que se van a realizar, puede procederse de manera iterativa: se sigue lanzando la aguja hasta que se considere que el error cometido en la estimación (la longitud del intervalo de confianza) es lo suficientemente pequeño.

3.3. Probabilidad de ganar un juego “solitario” de cartas

Supongamos que se desea determinar cuál es la probabilidad de ganar un juego “solitario” de cartas. Entendemos por un “solitario” un juego en el cual se emplea una baraja de 52 cartas y que tiene una estrategia fija.

Una aproximación razonable al problema es considerar que las 52! posibles ordenaciones de las cartas de la baraja son igualmente probables y calcular cuántas de ellas hacen que el jugador gane la partida.

Dado que no existe un método sistemático de determinar el número de combinaciones ganadoras más que ir probándolas una a una, y puesto que 52! es un número extraordinariamente grande, aparentemente el cálculo de la probabilidad de ganar el solitario parece un problema matemáticamente intratable.

Un método alternativo de abordar el problema es mediante simulación. Después de jugar n partidas, si definimos:

$$X_i = \begin{cases} 1 & \text{si la partida } i \text{ resulta en victoria} \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (3.12)$$

para $i : 1, \dots, n$, entonces X_i son variables aleatorias independientes de Bernoulli, para las cuales:

$$E\{X_i\} = \text{Prob}\{\text{Ganar la partida de “solitario”}\} \quad (3.13)$$

Por la *Ley de los Grandes Números*,

$$\frac{\sum_{i=1}^n X_i}{n} = \frac{\text{número de partidas ganadas}}{\text{número de partidas jugadas}} \quad (3.14)$$

convergerá con probabilidad 1 a

$$\text{Prob}\{\text{Ganar la partida de “solitario”}\} \quad (3.15)$$

Es decir, jugando un número elevado de veces, podemos usar la proporción de partidas ganadas para estimar la probabilidad de ganar el juego del solitario.

Para estudiar el problema del solitario, necesitamos diseñar un programa para jugar el juego a partir de una determinada ordenación de las cartas. La ordenación inicial de las cartas en cada partida debe ser aleatoria, y debemos obtenerla de modo que las 52! posibles ordenaciones sean equiprobables. A continuación, se describe un algoritmo que permite obtener la ordenación inicial aleatoria empleando únicamente números aleatorios.²

El algoritmo comienza escogiendo aleatoriamente una carta y poniéndola en la posición n . Seguidamente, escoge aleatoriamente entre las $n - 1$ cartas restantes y la pone en la posición $n - 1$; y así sucesivamente. Una forma eficiente de escoger entre las cartas restantes es mantener estos elementos en una lista ordenada y escoger la posición en la lista.

Para implementar el algoritmo es preciso generar una variable aleatoria uniforme discreta con rango $\{1, 2, \dots, k\}$. Se pueden obtener observaciones de esta variable aleatoria discreta a partir de números aleatorios u de la forma siguiente:

$$N_k = \lfloor k \cdot u \rfloor + 1 \quad (3.16)$$

donde $\lfloor k \cdot u \rfloor$ representa el mayor entero $\leq k \cdot u$.

²Puede encontrarse un algoritmo más eficiente para generar permutaciones de números enteros en (Ross 1994).

3.4. Cálculo de integrales definidas

Empleando la técnica de Monte Carlo, se pretende estimar el valor de una integral definida, $I = \int_a^b g(x) \cdot dx$, donde $g(x)$ es una función real monovaluada. Para ello, se define una variable aleatoria Y de la forma siguiente:

$$Y = (b - a) \cdot g(X) \quad (3.17)$$

La variable aleatoria X está distribuida uniformemente en el intervalo $[a, b]$. La densidad de probabilidad de X es:

$$f_X(x) = \frac{1}{b - a} \quad (3.18)$$

Como se demuestra a continuación, el valor esperado de la variable Y , μ_Y , es igual al valor de la integral definida que se desea calcular, I :

$$\begin{aligned} \mu_Y &= E\{(b - a) \cdot g(X)\} \\ &= (b - a) \cdot E\{g(X)\} \\ &= (b - a) \cdot \int_a^b g(x) \cdot f_X(x) \cdot dx \\ &= (b - a) \cdot \frac{\int_a^b g(x) \cdot dx}{(b - a)} \\ &= I \end{aligned} \quad (3.19)$$

Así pues, el problema de evaluar la integral definida I se ha reducido a estimar el valor esperado de la variable aleatoria Y . En particular, μ_Y puede estimarse calculando la media de cierto número n de observaciones independientes de Y , y_1, \dots, y_n , de la manera siguiente:

$$\bar{Y}(n) = \frac{\sum_{i=1}^n y_i}{n} = (b - a) \cdot \frac{\sum_{i=1}^n g(x_i)}{n} \quad (3.20)$$

donde x_i son observaciones independientes de una variable aleatoria X , que está distribuida uniformemente en el intervalo $[a, b]$. Puede demostrarse que (ver Apéndice A):

$$E\{\bar{Y}(n)\} = \mu_Y \quad (3.21)$$

$$\text{Var}\{\bar{Y}(n)\} = \frac{\sigma_Y^2}{n} \rightarrow \lim_{n \rightarrow \infty} \text{Var}\{\bar{Y}(n)\} = 0 \quad (3.22)$$

La técnica es extensible al cálculo de integrales del tipo:

$$I = \int_0^\infty g(x) \cdot dx \quad (3.23)$$

Para ello, puede realizarse el siguiente cambio de variable:

$$x_1 = \frac{1}{x + 1} \rightarrow dx_1 = \frac{-dx}{(x + 1)^2} = -x_1^2 \cdot dx \quad (3.24)$$

Sustituyendo en la integral, se obtiene:

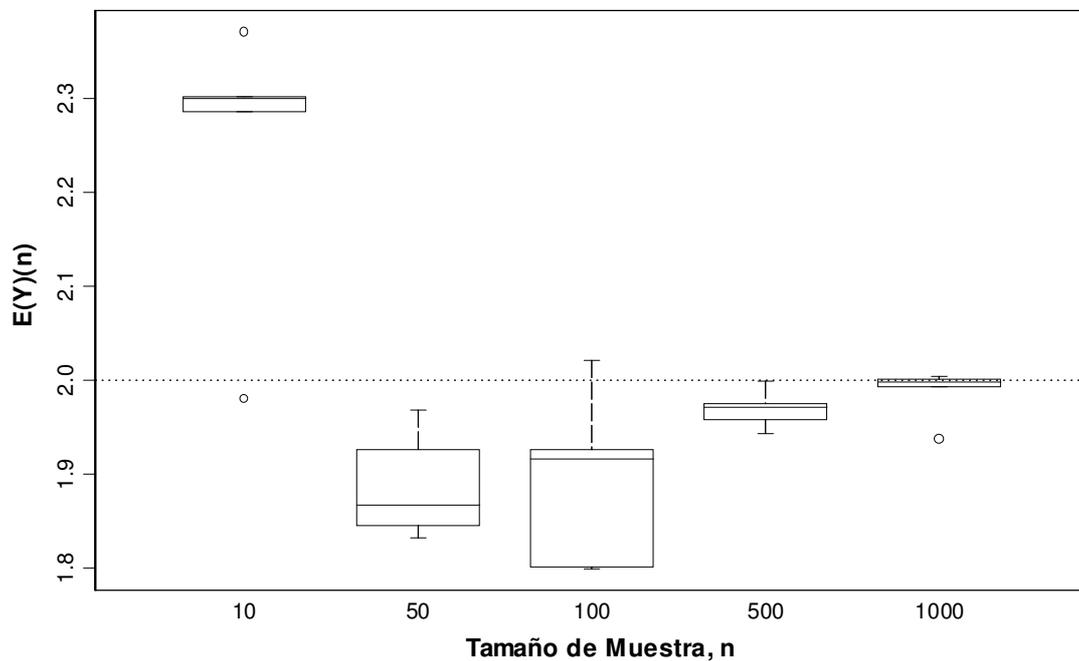


Figura 3.3: Boxplots de los resultados de la simulación.

$$I = \int_0^{\infty} g(x) \cdot dx = \int_0^1 \frac{g\left(\frac{1}{x_1} - 1\right)}{x_1^2} \cdot dx_1 \quad (3.25)$$

Ejemplo de aplicación

Como ejemplo, para ilustrar la aplicación de la técnica anteriormente descrita, supongamos que se desea estimar el valor de la integral definida siguiente:

$$I = \int_0^{\pi} \sin(x) \cdot dx \quad (3.26)$$

Se ha escogido una integral que posee primitiva analítica, $-\cos(x)$, a fin de comparar el resultado obtenido aplicando la técnica de Monte Carlo con el resultado analítico exacto, que es $I = 2$.

En el Anexo situado al final de este Tema, se muestra un programa escrito en C que realiza la simulación. La librería *stdlib.h*, disponible en el sistema UNIX y en ANSI C, contiene la función³, *int rand(void)*, destinada a la generación de números enteros uniformemente distribuidos en el intervalo $[0, 2^{15} - 1]$. Dividiendo *rand()* por 32767 se obtiene un número pseudo aleatorio (es decir, una observación de una variable aleatoria distribuida $U(0, 1)$).

A fin de analizar, por una parte cómo depende el valor estimado de I del tamaño de la muestra, n , y por otra parte la variabilidad de los resultados obtenidos, se realizan 5 réplicas

³Según se indica en el manual de referencia de la librería, el algoritmo de la función *rand* es un generador de números pseudo aleatorios multiplicativo congruente de periodo 2^{32} . Da como resultado un número entero comprendido entre cero y RAND_MAX. La constante simbólica RAND_MAX, definida en *stdlib.h*, vale $2^{15} - 1 = 32767$.

independientes de la simulación para cada uno de los siguientes 5 tamaños de muestra: 10, 50, 100, 500 y 1000. Los resultados de la media muestral obtenidos son los siguientes:

| | $n = 10$ | $n = 50$ | $n = 100$ | $n = 500$ | $n = 1000$ |
|------------------|----------|----------|-----------|-----------|------------|
| Réplica 1 | 1.980 | 1.832 | 2.021 | 1.958 | 2.001 |
| Réplica 2 | 2.302 | 1.845 | 1.799 | 1.971 | 1.993 |
| Réplica 3 | 2.286 | 1.926 | 1.801 | 1.999 | 1.998 |
| Réplica 4 | 2.371 | 1.867 | 1.926 | 1.943 | 1.937 |
| Réplica 5 | 2.300 | 1.968 | 1.916 | 1.975 | 2.004 |

La comparación visual entre los boxplots de las diferentes alternativas resulta muy explicativa⁴. En la Figura 3.3 se muestra la representación mediante boxplots de las cinco medias muestrales (una por réplica), para cada uno de los cinco valores de n . Al aumentar el tamaño de muestra los resultados tienden hacia el valor verdadero y su variabilidad disminuye. Una conclusión fundamental que debe extraerse de este ejemplo es que deben establecerse criterios para determinar el tamaño de muestra, n , que permita estimar I con una determinada precisión.

ANEXO

Programa de simulación para la estimación del valor de una integral definida

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>
#define PI 3.141593

struct intervalo {
    float inicio,
          fin;
};

float Uniforme_interv_A_B(interv)
struct intervalo interv;
{
    /* Generacion de U(A,B) */
    float v_unif;
    v_unif=rand()/32727.0;
    if (v_unif>1.0) v_unif=1.0;
    return(interv.inicio+(interv.fin-interv.inicio)*v_unif);
}

float Func_integrando(x)
float x;
{
    return(sin(x));
}

void Entrada_datos(num_muestras,interv_integ)
unsigned int      *num_muestras;
struct intervalo  *interv_integ;
{
    clrscr();
    /* Numero de muestras */
    printf("\tNumero de muestras (1-65535): ");
    scanf("%d",num_muestras);
    /* Intervalo de integracion */
    interv_integ->inicio=0;
    interv_integ->fin=PI;
}
```

⁴En el Apéndice A, al final del libro, se explica cómo se construye el boxplot de un conjunto de datos.

```
    /* Inicializacion rutina generacion numeros aleatorios */
    randomize();
    return;
}

void Informe(num_muestras,estima_integ)
unsigned int num_muestras;
float estima_integ;
{
    printf("\tI N F O R M E\n\t-----\n\n");
    printf("Numero de muestras: %d\n",num_muestras);
    printf("Estima de la integral: %f\n",estima_integ);
}

main()
{
    unsigned int    num_muestras,
                   contad=0;
    float           func_acumulada=0,
                   estima_integ;
    struct intervalo  interv_integ;
    Entrada_datos(&num_muestras,&interv_integ);
    while (contad<num_muestras) {
        func_acumulada+=Func_integrando(Uniforme_interv_A_B(interv_integ));
        contad++;
    }
    estima_integ=(interv_integ.fin-interv_integ.inicio)*
                 func_acumulada/num_muestras;
    Informe(num_muestras,estima_integ);
    return;
}
```


Tema 4

Modelado de sistemas de eventos discretos

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir y reconocer de qué partes se compone la especificación del modelo y del experimento.
- Discutir las diferentes estrategias para la simulación de modelos de eventos discretos.
- Discutir diferentes formalismos para el modelado de sistemas de eventos discretos.

4.1. Introducción

Básicamente, la especificación del modelo y del experimento consiste en definir:

- | | |
|-------------|--|
| MODELO | <ul style="list-style-type: none">• La <i>estructura lógica</i> y los <i>componentes</i> del modelo.• Los <i>aspectos matemáticos</i> del modelo. |
| EXPERIMENTO | <ul style="list-style-type: none">• Las <i>variables de entrada</i> y de <i>salida</i>.• Las <i>condiciones iniciales</i> del modelo.• La <i>condición de finalización</i> de la simulación.• Los <i>aspectos numéricos</i> de la simulación. |

La estructura lógica y los componentes del modelo se definen sobre la base de las *hipótesis de modelado* (o *aproximaciones de modelado*, como se prefiera). En función de la finalidad del estudio de simulación, se consideraran relevantes para ese propósito específico unos aspectos u otros del comportamiento del sistema, lo cual condiciona las hipótesis de modelado. El ámbito de validez de las hipótesis de modelado define el *marco experimental* del modelo.

Las variables de entrada y las condiciones iniciales determinan la forma en que el modelo es ejercitado durante la simulación. Además, la condición de finalización de la simulación y la elección de qué variables del modelo desean ser observadas (variables de salida), completan la definición de qué comportamientos del modelo desean ser estudiados en la simulación.

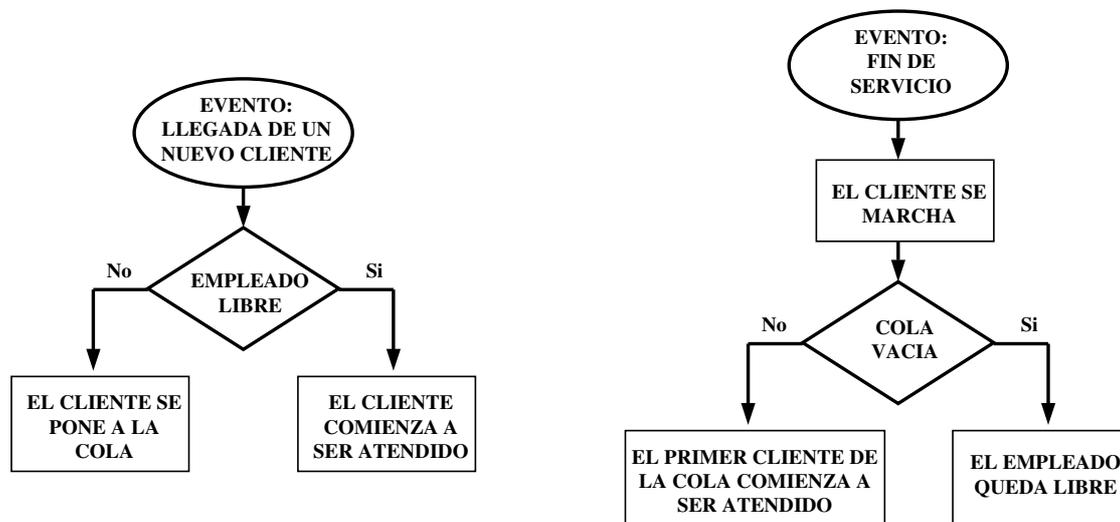


Figura 4.1: Acciones asociadas a cada uno de los dos tipos de evento.

Todo ello forma parte de la especificación del *experimento*. También es parte del experimento la especificación de qué métodos numéricos van a emplearse para simular el modelo.

4.2. Estructura lógica del modelo

A continuación se muestra un ejemplo de la definición de la *estructura lógica* de un modelo. Consideremos una oficina de atención al cliente en la que trabaja un único empleado:

- Si llega un nuevo cliente y el empleado está ocupado, el cliente se pone al final de una cola en espera de su turno. Si el empleado está libre, el cliente es atendido inmediatamente.
- Si el empleado termina de atender a un cliente, éste se marcha y comienza a ser atendido el primer cliente de la cola. Si la cola está vacía, el empleado permanece desocupado hasta la llegada de un nuevo cliente.

De su estructura lógica se deduce que en el modelo hay dos tipos de sucesos o eventos que desencadenan sendas secuencias de acciones (ver Figura 4.1). Estos dos eventos son:

- La llegada de un nuevo cliente a la oficina.
- La finalización de la atención a un cliente.

4.3. Elementos del modelo

Los modelos de tiempo discreto, en el contexto que nos ocupa, constan básicamente de los siguientes 7 tipos de componentes: las *entidades*, sus *atributos*, las *variables*, los *recursos*, las *colas*, los *contadores estadísticos* y los *eventos*. A continuación se describen brevemente.

Entidades

Las *entidades* son objetos dinámicos en la simulación, que son creados y se mueven por el sistema, cambiando el valor de sus atributos, afectados por otras entidades y por el estado del sistema. Las entidades pueden abandonar el sistema o bien permanecer indefinidamente circulando en él.

En el ejemplo anterior de la oficina, existe un único tipo de entidad: “cliente”. Cada cliente que se encuentra en la oficina es una *realización* (o *instanciación*, como se prefiera) de este tipo de entidad, que es creada a su llegada, circula a través del sistema y es destruida cuando el cliente abandona la oficina. Si en el modelo se hubieran definido diferentes tipos de clientes, que requirieran procesos de atención diferentes o a los que se asignara diferente prioridad, cada tipo de cliente se representaría como un tipo diferente de entidad.

Atributos

Los *atributos* permiten individualizar cada instanciación de una determinada clase de entidad. Al definir el tipo de entidad, se declaran sus atributos. Se pueden asignar valores diferentes a los atributos de cada instanciación de la clase de entidad, lo cual permite especificar las características particulares de cada uno de ellos.

Por ejemplo, algunos atributos que podrían definirse para el tipo de entidad “cliente” son: la prioridad con que debe ser atendido o determinados datos personales, como son el nombre y los apellidos, la edad, la nacionalidad, etc. En general, el valor de los atributos diferirá de un cliente a otro y es lo que permite diferenciarlos.

Variables

Las variables representan características del sistema que son independientes de los tipos de entidades o del número de realizaciones existentes en determinado instante. Por tanto, las variables no están asociadas a entidades en concreto, sino que pertenecen al conjunto del sistema. Son accesibles desde todas las entidades y pueden ser modificadas por todas las entidades. Puede considerarse que cada variable es como una pizarra colgada en la pared, en la que se escribe el valor de la variable. Todas las entidades pueden leer la pizarra, y también pueden borrar el valor escrito y escribir uno nuevo.

Algunas de las variables son intrínsecas a los elementos del modelo y, por ello surgen en casi todos los modelos de simulación. Algunas de estas son: el número de entidades (en nuestro caso, clientes) que hay en cada instante en cada cola, el número de recursos (en nuestro caso, empleados) ocupados, el estado (ocupado o libre) de cada recurso, el valor del reloj de la simulación (variable que va registrando el tiempo simulado) , etc.

Por el contrario, otras variables surgen debido a necesidades concretas del modelo en cuestión. Por ejemplo, supóngase que cada cliente tiene que ser atendido consecutivamente por dos empleados diferentes, situados a cierta distancia, y que el tiempo que emplea la entidad en ir de un empleado a otro se considera fijo. Entonces, el valor de este tiempo de “tránsito” sería una variable del modelo.

Las variables permiten *parametrizar* el modelo, es decir, definir magnitudes que pueden cambiar para adaptar el modelo a sus diferentes aplicaciones. Por ejemplo, mediante la modificación del tiempo de tránsito entre los dos puntos de atención, puede adaptarse un mismo modelo para representar diferentes sistemas.

Asimismo, las variables pueden representar magnitudes cuyo valor cambie durante el curso de la simulación. Por ejemplo: el número total de clientes que esperan, que están siendo atendidos y que se encuentran en tránsito entre los dos puntos de atención.

Recursos

Los *recursos* pueden ser el personal (en nuestro caso, el empleado), las máquinas (por ejemplo, si las entidades son piezas que deben ser procesadas), el espacio (por ejemplo, en un almacén), etc. Una entidad *captura* un recurso cuando éste está disponible, a fin de obtener un servicio de él, y lo *libera* una vez ha terminado.

El recurso puede ser individual o estar compuesto por un grupo de elementos individuales, cada uno de los cuales se llama una *unidad del recurso*. Por ejemplo, el caso de varios mostradores paralelos e idénticos de atención al público, puede representarse como un recurso con tantas *unidades* como puntos de atención. El número de unidades disponibles de un recurso puede variar durante el curso de la simulación, representando mostradores que son cerrados o abiertos.

Colas

Cuando una entidad no puede circular, debido tal vez a que necesita usar una unidad de un recurso que en ese momento no se encuentra disponible, entonces la entidad necesita un sitio donde esperar: este es el propósito de la *cola*.

Acumuladores estadísticos

A fin de calcular el valor de las variables de salida, es preciso calcular durante el curso de la simulación el valor de determinadas variables intermedias. Estas se llaman *acumuladores estadísticos*. Algunos ejemplos son: el número total de clientes atendidos hasta ese momento, la suma de los tiempos de espera en cola de los clientes hasta ese momento, el número total de clientes que han comenzado a ser atendidos hasta ese momento, el mayor tiempo de espera en cola hasta ese momento, etc. Los contadores son inicializados a cero al comenzar la simulación. Cuando “algo sucede” en la simulación (es decir, se ejecuta un evento), los contadores estadísticos afectados deben ser convenientemente actualizados.

Eventos

Un *evento* es un suceso que ocurre en un determinado instante de tiempo (simulado) y que puede cambiar el valor de los atributos, las variables y los acumuladores estadísticos.

Se denomina *actividad* al estado de una entidad en un determinado intervalo de tiempo.

En los modelos de tiempo discreto, los atributos, las variables y los acumuladores estadísticos sólo pueden cambiar a consecuencia de la ejecución de los eventos. Los cambios en sus valores se producen en los instantes en que son activados los eventos, manteniéndose constantes durante el intervalo de tiempo entre eventos sucesivos.

En nuestro ejemplo, hay cuatro tipos de eventos:

- Inicio de la simulación.
- Llegada a la oficina de un nuevo cliente.
- Fin de servicio a un cliente.
- Final de la simulación

Cada evento tiene asociado dos tipos de información:

- Su condición de activación, es decir, la condición que hace que el evento pase de estar desactivado a estar activado.
- Las acciones que deben realizarse en el instante en que el evento es activado.

En función del tipo de su condición de activación, los eventos pueden clasificarse en eventos en el tiempo y eventos en el estado:

- Los *eventos en el tiempo* son aquellos cuya condición de activación es que el tiempo de la simulación alcance un determinado valor. El valor del instante de tiempo simulado se almacena en una variable del modelo llamada *reloj de la simulación*.
- La condición de activación de los *eventos en el estado* no es función exclusiva del tiempo, sino que también es función de variables del sistema.

En los sistemas de tiempo discreto, dado que las variables y atributos permanecen constantes entre eventos consecutivos, los cambios en las variables necesarios para activar un evento en el estado sólo pueden producirse como resultado de la ejecución de un evento en el tiempo. La planificación de los eventos en el tiempo es muy sencilla, ya que se sabe con antelación en qué instantes se van a producir. Durante el curso de la simulación se lleva un registro de los instantes de activación de los eventos en el tiempo. Éste se denomina el *calendario de eventos*.

En los sistemas de tiempo discreto los cambios en las propiedades del sistema no se producen de manera continua en el tiempo. Por tanto, a fin de optimizar la simulación de este tipo de modelos, se hace avanzar el reloj de la simulación a saltos. Normalmente se usan dos métodos para ello: *avance hasta el siguiente evento* y *avance a incrementos de tiempo fijos*. Se volverá sobre este tema más adelante, al discutir las diferentes estrategias para la simulación de modelos de eventos discretos.

4.4. Aspectos matemáticos del modelo y las variables de entrada

Además de la estructura lógica y los componentes, deben especificarse los *aspectos matemáticos* del modelo, que están igualmente vinculados a las hipótesis de modelado.

En particular, debe decidirse qué comportamientos del modelo se consideran deterministas y cuáles aleatorios. Esta decisión condiciona qué variables de entrada son deterministas y cuáles aleatorias. En cada caso, la definición de la variable de entrada comprende:

- La dependencia temporal de las variables de entrada deterministas.
- La distribución de probabilidad de las variables de entrada aleatorias.

Continuando con el ejemplo, se consideran dos variables de entrada aleatorias:

- El tiempo que transcurre entre la llegada de un cliente y la llegada del siguiente cliente. Este tiempo, A , es una variable aleatoria que obedece a una distribución de probabilidad exponencial de media 1 minuto.
- El tiempo que tarda el empleado en atender a un cliente. Este tiempo, S , es una variable aleatoria exponencial con media 0.5 minutos.

En general, las distribuciones de probabilidad de las variables A y S deben determinarse, o bien estudiando los datos medidos del sistema real, o bien a través de consideraciones teóricas. En cualquier caso, el decidir que las variables de entrada están distribuidas de una determinada forma es otra hipótesis de modelado. El que el modelo reproduzca más o menos adecuadamente los aspectos de interés del sistema depende de lo acertado de estas aproximaciones.

Obsérvese que, al definir en el ejemplo las variables de entrada, se han realizado las hipótesis siguientes:

- Tanto la distribución de probabilidad de los intervalos entre llegadas, como la distribución de los tiempos de servicio, son independientes del instante de tiempo simulado. Sería fácil imaginar situaciones en las cuales durante determinadas “hora punta” la afluencia de clientes fuera mayor que durante las “horas valle”. También sería factible que, según avanza la jornada laboral, aumentara el cansancio del empleado, con lo cual aumentara el tiempo de servicio (o que disminuyera su paciencia, con lo cual el tiempo de servicio disminuiría). Pues bien, se ha realizado la hipótesis de que ninguno de estos factores, u otros similares, son relevantes.

- Los intervalos entre llegadas de los clientes y el tiempo de servicio son variables aleatorias estadísticamente independientes, es decir, no dependen de ninguna otra variable del modelo. Como en el caso anterior, es sencillo imaginarse situaciones en las cuales esto puede no ser así. Si el número de clientes esperando en la cola es elevado, el comportamiento del empleado en respuesta a la “presión” puede ser abreviar en lo posible los tiempos de atención. Con mayor motivo sucederá esto si se aproxima la hora de cierre de la oficina.
- Finalmente, se ha supuesto que ambas variables están distribuidas exponencialmente, y se ha fijado el valor de las medias.

Así pues, un aspecto extremadamente importante a la hora de modelar las variables de entrada es decidir acerca de su dependencia o independencia, tanto respecto a otras variables como respecto al tiempo. Si se sabe que el valor de una variable del sistema condiciona “en cierta medida” el valor de otra, entonces estas dos variables no son independientes estadísticamente. A la hora de generar observaciones de estas dos variables aleatorias durante la simulación, debe tenerse en cuenta esta dependencia, ya que de lo contrario los resultados del estudio pueden no ser válidos.

Decidir qué tipo de dependencias existen entre las variables aleatorias y describir estas dependencias adecuadamente es sin duda una de las tareas de modelado más delicadas y difíciles.

4.5. Medidas del comportamiento

Otra elección vinculada con el propósito del estudio es la de las *variables de salida*. Una variable de salida es aleatoria si y sólo si se calcula a partir de al menos una variable aleatoria. Cuando las variables aleatorias de salida miden el comportamiento del sistema a lo largo de toda la simulación, suele ser preciso definir en el modelo variables auxiliares, denominadas *acumuladores estadísticos* (o también *contadores estadísticos*), a partir de las cuales calcular la variable de salida. A continuación se muestran algunos ejemplos:

- El *tiempo medio de espera del cliente en la cola*, se define de la forma siguiente:

$$\hat{d}(n) = \frac{\sum_{i:1}^n D_i}{n} \quad (4.1)$$

donde n es el número total del clientes que han abandonado la cola y D_i es el tiempo de espera en la cola del cliente i .

Para calcularlo es preciso llevar registro, a lo largo de la simulación, de:

- El número de clientes que han abandonado la cola hasta ese momento, n .
- La suma del tiempo de espera de los clientes que han abandonado la cola hasta ese momento: $D(n) = \sum_{i:1}^n D_i$.

- El *número medio de clientes que componen la cola* se define de la forma siguiente:

$$\hat{q}(T) = \frac{\int_0^T Q(\tau) \cdot d\tau}{T} \quad (4.2)$$

donde T es el tiempo que dura la simulación y $Q(\tau)$ es el número de clientes que hay en la cola en el instante τ .

Para calcularlo, es preciso llevar registro a lo largo de la simulación de:

- El valor de la integral: $\int_0^t Q(\tau) \cdot d\tau$. Para ello se define un acumulador estadístico, $R(t)$. Inicialmente R vale cero. Cada vez que cambia el número de clientes de la cola, debe actualizarse el valor de R : se suma al valor actual de R el producto del número de clientes de la cola (antes del cambio) por el tiempo transcurrido desde el

anterior cambio. De este modo va calculándose el área bajo la curva $Q(t)$. Para ello, es preciso definir otro acumulador estadístico, t_{evento} , que almacene el instante en que se produjo el anterior evento.

Obsérvese que hay dos tipos de eventos que pueden modificar el tamaño de la cola: la llegada de un cliente y el final del servicio a un cliente (el primero de la cola comienza entonces a ser atendido). Si la condición de finalización no coincide con uno de estos dos tipos de eventos, al término de la simulación debe también actualizarse R .

- Para calcular el *número total de clientes atendidos*, es preciso llevar registro durante la simulación de:
 - El número de clientes que han sido atendidos hasta el momento.
- Para calcular el *tiempo máximo de espera en la cola* o el *número máximo de clientes de la cola*, es preciso conocer durante la simulación, respectivamente, cual es:
 - El tiempo de espera y el número de clientes de la cola máximo hasta ese momento.
- El *porcentaje del tiempo que el empleado está ocupado* puede calcularse de la forma siguiente:

$$\hat{f}_{\text{ocupación}}(T) = \frac{\int_0^T f_{\text{ocupación}}(\tau) \cdot d\tau}{T} \quad (4.3)$$

donde el factor de ocupación $\hat{f}_{\text{ocupación}}$ vale uno cuando el empleado está ocupado y cero cuando está libre. Para poder realizar el cálculo, es preciso llevar registro, a lo largo de la simulación, de:

- El valor de la integral: $\int_0^T f_{\text{ocupación}}(\tau) \cdot d\tau$. Para ello, se define un acumulador que inicialmente vale cero. Cada vez que el empleado cambia de ocupado a libre se suma al acumulador el tiempo que ha permanecido ocupado. Para ello debe definirse otro acumulador que almacene el instante anterior en que el empleado pasó de estar libre a ocupado.

Normalmente existen muchas posibles medidas del comportamiento del sistema, y probablemente es una buena idea observar tantas como sea (razonablemente) posible, ya que una vez analizados los resultados es posible ignorar algunos de ellos, sin embargo, no es posible analizar datos de los que uno no dispone (¡y en algunos casos uno puede llevarse alguna sorpresa!). La única desventaja de ampliar la recogida de datos es que se penaliza el tiempo de ejecución de la simulación.

4.6. Condiciones inicial y final de la simulación

La especificación de las condiciones de *inicio* y de *finalización* de la simulación es una parte obligada en la definición del experimento. Es preciso asignar valor inicial únicamente a aquellas variables del sistema cuyo valor futuro dependa de su valor presente. Estas variables se denominan *variables de estado*. Una propiedad de las variables de estado es que su valor actual representa completamente la “historia” del sistema, es decir, el comportamiento futuro del sistema puede determinarse únicamente conociendo el valor actual de sus variables de estado y el valor presente y futuro de las variables de entrada.

Un ejemplo comúnmente empleado para ilustrar el concepto de variable de estado es el de una partida de ajedrez. Para poder reestablecer la partida desde cualquier jugada es suficiente con conocer la posición de las fichas en ese momento, no siendo necesario conocer la secuencia de movimientos que ha conducido a esa situación. Por tanto, la posición de las fichas describe completamente el estado del sistema.

En nuestro ejemplo existen dos variables cuyo valor futuro depende del valor presente:

- El estado del empleado: E . Toma dos valores: libre ($E = 0$) u ocupado ($E = 1$).
- El número de clientes que esperan en la cola: $Q(t)$.

Además, por supuesto, de los acumuladores estadísticos, que deben ser inicializados convenientemente.

Una posible condición inicial es que el empleado se encuentre libre y la cola vacía. Sin embargo, son imaginables otras muchas alternativas. Supóngase, por ejemplo, que los clientes empiezan a llegar antes de que se abra la oficina al público, de modo que cuando ésta comienza sus funciones el tamaño de la cola no es cero. Es más, si el número de clientes de la cola en el instante inicial fluctúa significativamente de unos días a otros, éste deberá considerarse una variable aleatoria. En este caso, con cada réplica de la simulación habría que generar una observación de la variable aleatoria “tamaño inicial de la cola”. También son posibles otros enfoques, por ejemplo, podría obtenerse el tamaño inicial de la cola modelando la frecuencia de llegada de los clientes mientras la oficina está cerrada. Todo ello depende, evidentemente, de las características del sistema y de las hipótesis de modelado.

Lo mismo sucede con la condición de finalización de la simulación. Esta condición puede depender únicamente del tiempo simulado (por ejemplo, que el tiempo simulado sea superior a 8 horas), de que alguna de las variables del modelo alcance determinado valor o de la satisfacción de ambos tipos de condiciones (en un esquema más “razonable” de atención al público, que el tiempo sea superior a 8 horas y no queden clientes por atender).

4.7. Descripción del funcionamiento del sistema

Llegado este punto, se han descrito:

- los componentes del sistema (entidades, recursos y colas),
- las variables que los describen (atributos, variables y contadores estadísticos), y
- la interacción entre ellos (eventos).

No obstante, todavía falta describir en el modelo los detalles acerca del funcionamiento del sistema. Básicamente, puede realizarse desde dos ópticas distintas:

- la *orientación a los eventos*, y
- la *orientación a los procesos*.

Modelado orientado a los eventos

Como su nombre indica, el *modelado orientado a los eventos* se centra entorno a la descripción de los eventos, es decir:

- qué tipos de eventos se producen,
- qué condición de activación tiene cada uno, y
- cuál es el flujo lógico de acciones asociadas a la activación de cada evento.

En la Figura 4.2 se muestran los flujos de acciones correspondientes a los eventos del modelo de la oficina de atención al público: inicialización, llegada de un nuevo cliente; y marcha de un cliente. La condición de finalización de la simulación es que el número de clientes que comienzan a ser atendidos, n , sea superior a uno establecido. Por simplicidad, se supone en este ejemplo que las variables de salida son únicamente:

- El tiempo medio de espera en la cola: $\hat{d}(n) = \frac{\sum_{i=1}^n D_i}{n} = \frac{D(n)}{n}$.
- El número medio de clientes que componen la cola: $\hat{q}(T) = \frac{\int_0^T Q(\tau) \cdot d\tau}{T} = \frac{R(T)}{T}$.

Dos de las ventajas fundamentales del modelado orientado a los eventos son:

- Permite una flexibilidad total en la descripción del modelo, y
- La realización, empleando un lenguaje de programación, del código de la simulación a partir de este tipo de descripción del modelo es conceptualmente sencilla.

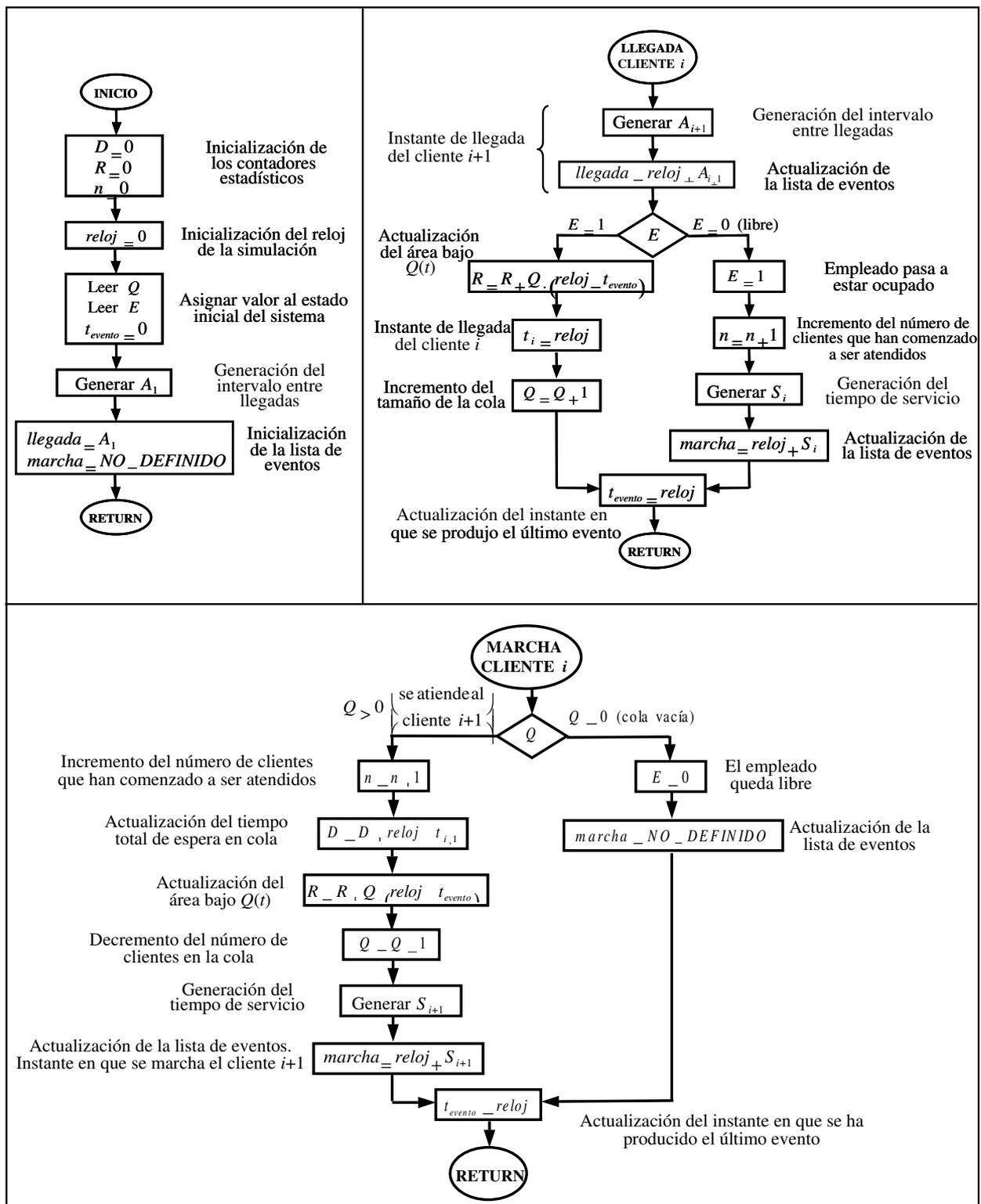


Figura 4.2: Flujos de acciones asociadas a los eventos.

Sin embargo, la orientación a los eventos presenta también una desventaja importante: la realización de modelos de grandes dimensiones, con diferentes tipos de eventos, entidades y recursos, resulta excesivamente compleja. Este enfoque requiere que el programador adquiera el papel de “supervisor omnisapiente”, llevando el control de todos los eventos, entidades, atributos, variables y acumuladores estadísticos.

Modelado orientado a los procesos

Una forma alternativa, más natural y sencilla, de describir el modelo consiste en tomar el punto de vista de las entidades y describir su circulación a través del sistema. Este enfoque se centra en los procesos que llevan a cabo las entidades, por ello se llama *modelado orientado a los procesos*. Su práctica es posible gracias al empleo de lenguajes de simulación, que traducen de manera automática la descripción orientada a los procesos a una descripción orientada a los eventos, y ésta en código escrito en algún lenguaje de programación. En última instancia, el código ejecutable de la simulación siempre está orientado a los eventos.

El modelo orientado a los procesos de la oficina de atención al público se realiza tomando el punto de vista de un cliente cualquiera. Como en el caso anterior, las variables de salida son el tiempo medio de espera en la cola y el número medio de clientes que componen la cola. Los pasos en el proceso de atención son:

- Llego a la oficina.
- Escribo en mi atributo “Instante de llegada” el valor que tiene en este momento el reloj de la simulación. Así más tarde podré calcular el tiempo que he estado esperando en la cola.
- Me pongo al final de la cola e incremento en uno el valor de la variable “Número de clientes de la cola”.
- Espero hasta que yo sea el primero de la cola y el empleado esté libre (si tengo suerte, el tiempo de espera será cero).
- En el instante en que abandono la cola, calculo mi tiempo de espera (restando el valor de mi atributo “Instante de llegada” del valor del reloj de la simulación), decremento en uno el valor de la variable “Número de clientes de la cola”, incremento en uno la variable “Número de clientes que abandonan la cola” y comienzo a ser atendido por el empleado.
- El empleado me atiende durante el tiempo que requiero.
- Finaliza mi tiempo de servicio, con lo que dejo libre al empleado y abandono la oficina.

En los dos próximos temas se abordan en detalle las dos metodologías de modelado anteriormente citadas: el modelado orientado a los eventos y el modelado orientado a los procesos.

EJERCICIOS

4.1 Describa, empleando la metodología de la *orientación a los eventos*, el modelo que usted ha propuesto al contestar al Problema 2.1. En particular, responda a las cuestiones siguientes:

- Cuáles son los eventos.
- Cuál es la condición de activación de cada evento.
- Cuáles son las acciones asociadas a cada evento.
- Indique cuáles son las condiciones inicial y final de la simulación.

4.2 Describa, empleando la metodología de la *orientación a los procesos*, el modelo que usted ha propuesto al contestar al Problema 2.1. En particular, responda a las cuestiones siguientes:

- Cuáles son las entidades.
- Cuáles son procesos. Cuál es el recurso de cada proceso.
- En cada proceso, qué acción ejerce la entidad sobre el recurso.
- Describa el flujo de las entidades a través del modelo.

Tema 5

Simulación usando un lenguaje de programación

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los fundamentos del modelado orientado a los eventos.
- Diseñar y realizar evaluaciones básicas de modelos orientados a los eventos.
- Diseñar programas de simulación de modelos orientados a los eventos, para su codificación mediante un lenguaje de programación de propósito general.
- Discutir las características de los sistemas de colas y de inventario.
- Modelar sistemas sencillos empleando la metodología de los grafos de eventos.

NOTA IMPORTANTE: En los Anexos 1 y 2, situados al final del Tema, se muestra el código C de dos modelos: una cola atendida por un empleado y la gestión de un inventario. Se trata de material complementario, que el alumno puede consultar o no según juzgue más conveniente. En el examen no se exigirá el conocimiento de ningún lenguaje de programación en particular, ya que no se pedirá al alumno que genere código, sino diagramas de flujo similares a los mostrados a lo largo del Tema.

5.1. Introducción

En este tema se va a discutir el diseño de modelos de tiempo discreto para su simulación empleando un lenguaje de programación. Esta metodología de modelado recibe el nombre de *modelado orientado a los eventos*, ya que la descripción del modelo básicamente consiste en definir:

- Qué condiciones deben satisfacerse para que se active cada tipo de evento.
- Qué acciones deben realizarse cuando se activa cada tipo de evento.

La orientación a los eventos es una metodología que hoy en día se encuentra prácticamente en desuso, ya que resulta más eficiente el empleo de lenguajes de simulación y entornos de modelado que soporten el *modelado orientado a los procesos*.

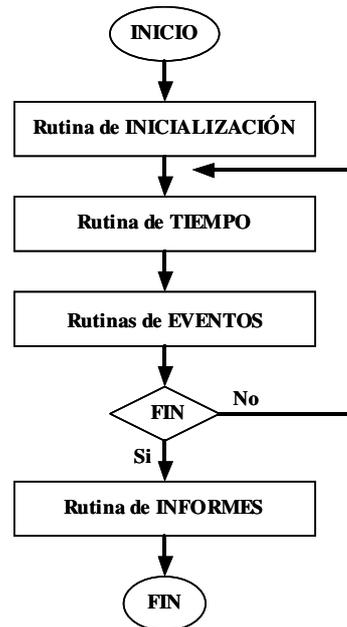


Figura 5.1: Diagrama de flujo de la simulación de un modelo de tiempo discreto.

Sin embargo, el interés pedagógico del conocimiento de esta metodología es innegable: una vez que se aprende a programar las simulaciones usando un lenguaje de programación, debiéndose prestar atención hasta a los detalles más mínimos, es menos probable que se comenten errores conceptuales cuando se empleen más adelante los lenguajes de simulación.

En las siguientes secciones se discute el algoritmo de simulación de modelos de tiempo discreto empleando la metodología de la orientación a los eventos, y se aplica al modelado de dos sistemas: una cola atendida por un único empleado y la gestión de inventario de un único producto. Se ha usado en ambos casos el lenguaje de programación C: al permitir gestionar dinámicamente la memoria hace más elegante y eficiente la gestión de las colas. En los dos Anexos situados al final del tema se muestra el código fuente de ambos modelos de simulación.

5.2. Algoritmo de simulación de modelos de tiempo discreto

Aunque existen diversas estrategias, la mayor parte de los modelos de simulación están esencialmente compuestos por las rutinas siguientes:

- *Rutina de inicialización.* Asigna valores iniciales a las variables de estado e inicializa los contadores estadísticos.
- *Rutina de tiempo.* Determina cuál es el siguiente evento (el más inminente), de entre los del calendario de eventos, y avanza el reloj de la simulación hasta ese instante.
- *Rutinas de eventos.* Son las rutinas, una para cada tipo de evento, que realizan el flujo de acciones asociado al evento. Actualizan el valor de las variables, los atributos y contadores estadísticos, así como el calendario de eventos.
- *Rutina de informes.* Al finalizar la simulación, calcula y muestra el valor de las variables de salida.

El *código ejecutivo* controla el flujo de control de la simulación, tal como se muestra en la Figura 5.1. Al iniciarse la simulación, el código ejecutivo pasa el control a la *rutina de inicialización*, para que ésta inicialice el reloj de la simulación, las variables de estado, la lista de eventos y los acumuladores estadísticos.

Una vez ejecutadas las acciones de inicialización, el código ejecutivo transfiere el control a la *rutina de tiempo*, para que avance el reloj de la simulación y evalúe el tipo de evento a ejecutar. Pueden emplearse dos mecanismos de avance en el tiempo: *avance a incrementos fijos* o *avance hasta el evento más próximo*.

A continuación, el código ejecutivo transfiere el control a la rutina correspondiente al tipo de evento activado. Las acciones realizadas por la *rutina de evento* dependen de cada caso, pero en general la rutina actualizará las variables de estado, los contadores estadísticos y añadirá nuevos eventos a la lista de eventos.

Existe un evento especial, el evento de *finalización de la simulación*, que se activa cuando se satisfacen las condiciones fijadas para la finalización de la simulación, por ejemplo, que se alcance determinado valor del tiempo, que el estado del sistema satisfaga determinadas condiciones, etc. Cuando se activa el evento de finalización, el código ejecutivo transfiere el control a la *rutina generadora de informes* y una vez concluidas las tareas de ésta, finaliza la simulación.

Si el evento ejecutado no es el de finalización, el código ejecutivo devuelve el control a la *rutina de tiempo*.

Mecanismos de avance del reloj de la simulación

En los modelos de tiempo discreto, los cambios en el estado del sistema físico se representan mediante una serie de cambios discretos o *eventos* en instantes específicos de tiempo. Entre eventos, las variables de estado del sistema permanecen constantes. A consecuencia de ello, la actualización de reloj de la simulación al siguiente instante de tiempo se realiza sin consultar ni modificar ninguna variable de estado.

Pueden emplearse dos métodos para avanzar el reloj de la simulación. El primer método es la *simulación orientada a los intervalos*, donde el reloj de la simulación es avanzado a incrementos de tiempo fijos, Δt . El segundo método es la *simulación orientada a los eventos*, donde el reloj se incrementa hasta el instante en el cual se activa el evento más reciente en el tiempo. A continuación, se describe brevemente cada uno de ellos.

Avance en incrementos fijos de tiempo: *simulación orientada a los intervalos.* El reloj de la simulación avanza en incrementos de tiempo fijos convenientemente escogidos, Δt . Una vez ejecutado en el instante inicial de la simulación, t_{inicial} , el evento de inicialización, y actualizado correspondientemente el calendario de eventos, el reloj de la simulación salta al instante $t_{\text{inicial}} + \Delta t$. Entonces se comprueba, consultando el calendario, si en el intervalo de tiempo $(t_{\text{inicial}}, t_{\text{inicial}} + \Delta t]$ estaba planificado algún evento. Si es así, se ejecuta suponiendo que ha sido activado en $t_{\text{inicial}} + \Delta t$. Es decir, se considera (de manera aproximada) que todos los eventos ocurridos en el intervalo han sucedido al final del mismo, actualizándose el estado del sistema en consecuencia. Una vez ejecutados los eventos y actualizado (si procede) el calendario, el reloj de la simulación salta al instante $t_{\text{inicial}} + 2 \cdot \Delta t$ y así sucesivamente hasta que se ejecuta el evento de finalización.

El error cometido disminuye al hacer más pequeño Δt , sin embargo, esto conlleva un aumento en el número de intervalos y, por tanto, en el número de comprobaciones de si se han producido o no eventos, lo cual supone un incremento en el coste computacional de la simulación. Es preciso, por ello, establecer un balance entre error y coste computacional.

Este método resulta idóneo cuando se sabe a priori que los eventos suceden siempre con una frecuencia fija conocida. Por ejemplo, si los datos de un sistema económico suceden con una frecuencia anual, parece natural hacer avanzar el reloj de la simulación a incrementos de tiempo, Δt , de un año.

Avance hasta el siguiente evento: *simulación orientada a los eventos*. En el instante de inicio de la simulación, se activa el evento “Inicio de la Simulación”. Como parte de las acciones asociadas a la ejecución de este evento, se pone el reloj de la simulación a cero y se planifican determinados eventos para su ejecución en instantes futuros. Estos eventos en el tiempo son registrados en el *calendario de eventos*, ordenados de menor a mayor instante de ejecución. Una vez ejecutado este primer evento, el reloj de la simulación es avanzado hasta el instante de ejecución del primer evento del calendario, el cual es entonces borrado del calendario. Se ejecuta el evento, actualizándose, si procede, el calendario de eventos. A continuación, el reloj de la simulación salta hasta el instante de ejecución del evento más próximo (el primero del calendario), éste es borrado del calendario, es ejecutado y, si procede, se actualiza el calendario. Se procede de esta manera hasta que se activa el evento “Finalización de la Simulación”. Una vez éste es ejecutado, finaliza la simulación.

5.3. Simulación de una cola atendida por un empleado

Consideremos de nuevo el modelo de la oficina de atención al cliente planteado en el tema anterior, y veamos qué aspectos hay que considerar para programar el modelo empleando el lenguaje C.

Generación de observaciones de una variable aleatoria exponencial

Las dos variables de entrada del modelo (intervalo entre llegadas y tiempo de servicio) son aleatorias, con probabilidad distribuida exponencialmente. Pueden obtenerse observaciones de una variable aleatoria exponencial, Y , partir de números *seudo aleatorios* (es decir, observaciones de una variable aleatoria uniforme), X , mediante la transformación siguiente¹:

$$y = -m \cdot \ln(x) \quad (5.1)$$

Esto puede expresarse, de manera algo más formal, del modo siguiente:

$$\left. \begin{array}{l} X \sim U(0, 1) \\ y = -m \cdot \ln(x) \end{array} \right| \rightarrow Y \sim \text{expo}(m) \quad (5.2)$$

La librería *stdlib.h*, disponible en el sistema UNIX y en ANSI C, contiene la función *int rand(void)*, destinada a la generación de números enteros uniformemente distribuidos en el intervalo $[0, 2^{15} - 1]$. Dividiendo la salida de *rand()* por 32767, se obtiene un número seudo aleatorio distribuido $U(0, 1)$. En el Anexo 1, situado al final del Tema, se muestra el programa de simulación. La función *Distr_expon* devuelve una observación de una variable distribuida exponencialmente. El valor de la media de la distribución, m , es el argumento de la función.

Gestión de las colas

Se han programado las cuatro funciones siguientes para la gestión de las colas (ver el fichero *cola.h*, en el Anexo 1):

- *Esta_vacia*. Informa acerca de si la cola posee algún elemento.
- *Anadir_elemento*. Añade un nuevo elemento al final de la cola.

¹**Demostración.** Sea una variable aleatoria uniforme, $X \sim U(0, 1)$, y otra variable aleatoria Y definida a partir de X de la forma siguiente: $Y = -m \cdot \ln(X)$. La función de distribución acumulada de la variable aleatoria Y puede calcularse de la forma siguiente:

$$F_Y(y) = \text{Prob}\{Y \leq y\} = \text{Prob}\{-m \cdot \ln(X) \leq y\} = \text{Prob}\left\{X \geq e^{-\frac{y}{m}}\right\} = 1 - e^{-\frac{y}{m}}$$

Esta función de distribución acumulada, $F_Y(y) = 1 - e^{-\frac{y}{m}}$, es la de una variable *distribuida exponencialmente*, así pues, queda demostrado que $Y \sim \text{expo}(m)$, es decir, que Y está distribuida exponencialmente con media m .

Tabla 5.1: Nombre y finalidad de las rutinas programadas.

| PROPÓSITO | NOMBRE DE LA RUTINA | ACCIONES |
|---|--|---|
| Inicialización | <i>Inicializacion</i> | Inicializa el reloj de la simulación, las variables de estado, los contadores estadísticos, el calendario de eventos y el generador de números aleatorios. |
| Tiempo | <i>Rutina_tiempo</i> | Avanza el reloj de la simulación hasta el evento más inminente de los del calendario de eventos. |
| Eventos | <i>Evento_llegada</i> <i>Evento_marcha</i> | Actualiza las variables de estado, el calendario de eventos y los contadores estadísticos. Actualiza las variables de estado, el calendario de eventos y los contadores estadísticos. |
| Decisión sobre la finalización de la simulación | <i>Final_simulacion</i> | Comprueba si se satisface alguna de las condiciones de finalización de la simulación. |
| I/O | <i>Entrada_datos</i> <i>Informe</i> <i>Imprime_estado_simulacion</i> | Entrada por teclado de los valores medios de las distribuciones exponenciales y de la condición de finalización. Calcula las variables de salida y muestra sus valores. Muestra el valor actual del reloj de la simulación, de las variables de estado, los contadores estadísticos y el calendario de eventos. |
| Generación de observaciones de variables aleatorias | <i>Distr_expon</i> | Genera observaciones de una variable aleatoria distribuida expo (m). |

- *Sacar_elemento*. Extrae el primer elemento de la cola.
- *Imprime_cola*. Realiza un listado del contenido de la cola.

Rutinas y programa principal

La condición de finalización de la simulación es que el número de clientes que comienzan a ser atendidos alcance determinado valor previamente establecido. Este número, junto con la media de las distribuciones estadísticas de las dos variables de entrada, son los tres datos de entrada al programa.

La salida del programa es el valor de las variables de estado, del calendario de eventos (o lista de eventos) y de los contadores estadísticos, cada vez que se produce un evento. Además, al finalizar la simulación, debe mostrarse el valor de las dos variables de salida: el tiempo medio de espera en la cola y el número medio de clientes en la cola.

La Tabla 5.1 contiene una explicación de la funcionalidad de cada una de las rutinas programadas (ver el Anexo 1). Los diagramas de flujo de la rutina de inicialización y de las rutinas de eventos son las mostradas en la Figura 4.2.

El flujo del programa principal es el mostrado en la Figura 5.1. Llama a las rutinas de inicialización y de entrada de datos. Mientras la rutina de finalización no indique que la simulación ha concluido, llama sucesivamente a la rutina de tiempo y a la rutina del evento correspondiente en cada caso. Al final de la simulación, llama a la rutina de informe.

El estudio de simulación

A continuación se muestra un ejemplo de la evolución del sistema durante la simulación. El tiempo está medido en minutos. Los valores de entrada de la simulación son los siguientes:

Entrada de datos:

```
Media del intervalo entre llegadas:          1.0
Media del tiempo de servicio:              0.5
Num. clientes cuyo tiempo de espera debe   5
simularse (1-65535):
```

Obsérvese que la condición de finalización es que el número de clientes que comienzan a ser atendidos, n , sea igual a cinco. La salida del programa consiste en un listado de:

- el valor de las variables del sistema y los contadores estadísticos cada vez que se ejecuta un evento, y
- una vez finalizada la simulación, el informe con los valores de las variables de salida.

A continuación se muestran los resultados obtenidos de realizar una réplica de la simulación, con las condiciones iniciales y de finalización anteriormente indicadas. En la Figura 5.2 se representa la evolución del sistema. Los círculos representan a los clientes y el rectángulo al empleado. Mientras un cliente está siendo atendido, el círculo se dibuja inscrito en el rectángulo. Mientras el empleado está libre, el rectángulo no contiene ningún círculo. El número dentro del círculo es el instante en que ha llegado el cliente a la cola.

```
----- Reloj: 0.000000
Empleado LIBRE  0 clientes en la cola  Ultimo evento: 0.000000
Cola: Cola vacia
Lista de eventos:  Llegada = 4.549548  Marcha = -
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 0
  Tiempo total de espera de los clientes en la cola:  0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.549548
Empleado OCUPADO  0 clientes en la cola  Ultimo evento: 4.549548
Cola: Cola vacia
Lista de eventos:  Llegada = 6.371961  Marcha = 4.796673
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 1
  Tiempo total de espera de los clientes en la cola:  0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.796673
Empleado LIBRE  0 clientes en la cola  Ultimo evento: 4.796673
Cola: Cola vacia
Lista de eventos:  Llegada = 6.371961  Marcha = -
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 1
  Tiempo total de espera de los clientes en la cola:  0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 6.371961
Empleado OCUPADO  0 clientes en la cola  Ultimo evento: 6.371961
Cola: Cola vacia
Lista de eventos:  Llegada = 6.494414  Marcha = 6.453289
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola:  0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 6.453289
Empleado LIBRE  0 clientes en la cola  Ultimo evento: 6.453289
Cola: Cola vacia
Lista de eventos:  Llegada = 6.494414  Marcha = -
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola:  0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 6.494414
Empleado OCUPADO  0 clientes en la cola  Ultimo evento: 6.494414
Cola: Cola vacia
```

```

Lista de eventos:  Llegada = 7.640840  Marcha = 7.127680
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 3
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 7.127680
Empleado LIBRE 0 clientes en la cola Ultimo evento: 7.127680
Cola: Cola vacia
Lista de eventos:  Llegada = 7.640840  Marcha = -
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 3
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 7.640840
Empleado OCUPADO 0 clientes en la cola Ultimo evento: 7.640840
Cola: Cola vacia
Lista de eventos:  Llegada = 8.381031  Marcha = 8.910286
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 4
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 8.381031
Empleado OCUPADO 1 clientes en la cola Ultimo evento: 8.381031
Cola: 8.381031
Lista de eventos:  Llegada = 8.597755  Marcha = 8.910286
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 4
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 8.597755
Empleado OCUPADO 2 clientes en la cola Ultimo evento: 8.597755
Cola: 8.381031 8.597755
Lista de eventos:  Llegada = 12.388444  Marcha = 8.910286
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 4
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.216724

----- Reloj: 8.910286
Empleado OCUPADO 1 clientes en la cola Ultimo evento: 8.910286
Cola: 8.597755
Lista de eventos:  Llegada = 12.388444  Marcha = 9.028479
Cont. estadísticos:
  Numero de clientes que han empezado a ser atendidos: 5
  Tiempo total de espera de los clientes en la cola: 0.529255
  Integral num. clientes en la cola respecto del tiempo: 0.841785

```

Fin de la simulacion...

I N F O R M E

```

-----
Tiempo medio de espera en la cola:          0.105851
Numero medio de clientes en la cola:        0.094473

```

En esta réplica de la simulación, sólo un cliente abandona la cola: entra en la cola en el instante 8.381031 minutos y la abandona, para ser atendido, en el instante 8.910286 minutos, que coincide con el final de la simulación. Al finalizar la simulación hay un cliente esperando en la cola. Como la simulación finaliza antes de que éste abandone la cola, este cliente no influye en el valor de los contadores estadísticos.

El tiempo medio de espera en la cola se calcula dividiendo el tiempo total de espera,

$$8.910286 - 8.381031 = 0.529255 \text{ minutos}$$

entre el número total de clientes que han abandonado la cola, $n = 5$:

$$\hat{d}(5) = \frac{0.529255}{5} = 0.105851 \text{ minutos/cliente} \quad (5.3)$$

El número medio de clientes en la cola se calcula integrando en el tiempo el número de clientes a la cola y dividiendo por el tiempo simulado:

- Entre los instantes 0 y 8.381031 minutos la cola ha estado vacía, con lo cual ese intervalo no contribuye a la integral.
- Entre 8.381031 minutos y 8.597755 minutos había 1 cliente en la cola, con lo cual, la contribución a la integral es:

$$(1 \text{ cliente}) \cdot (8.597755 \text{ minutos} - 8.381031 \text{ minutos}) = 0.216724 \text{ clientes} \cdot \text{minuto}$$

- Entre los instantes 8.597755 minutos y 8.910286 minutos había 2 clientes en la cola, con lo cual la contribución a la integral de este intervalo es:

$$(2 \text{ clientes}) \cdot (8.910286 \text{ minutos} - 8.597755 \text{ minutos}) = 0.625062 \text{ clientes} \cdot \text{minuto}$$

El valor total de la integral es:

$$R(8.910286) = 0.216724 + 0.625062 = 0.841786 \text{ clientes} \cdot \text{minutos} \quad (5.4)$$

Dividiendo por el tiempo de simulación, 8.910286 minutos, se obtiene el número medio de clientes en la cola:

$$\hat{q}(8.910286) = \frac{R(8.910286)}{8.910286} = \frac{0.841786 \text{ clientes} \cdot \text{minutos}}{8.910286 \text{ minutos}} = 0.094473 \text{ clientes} \quad (5.5)$$

Los resultados de una ejecución de la simulación son conceptualmente análogos a un conjunto de observaciones del comportamiento del sistema (correspondientes al intervalo de tiempo necesario para que 5 clientes comiencen a ser atendidos). Si el instante en que comienzan a recogerse los datos del sistema varía, por ejemplo, de un día a otro, los intervalos entre llegadas y de los tiempos de servicio observados serán significativamente diferentes y, en consecuencia, el valor calculado de los estadísticos que representan el comportamiento del sistema (tiempo de espera medio y tamaño medio de la cola) será diferente. Los estadísticos son variables aleatorias. El valor de cada uno, obtenido de realizar una única simulación, es una “muestra de tamaño uno”.

Para poder extraer conclusiones acerca de cómo se comporta “típicamente” el sistema (no en un momento en concreto, sino en general), es necesario repetir varias veces el mismo experimento. Igualmente, para poder extraer conclusiones acerca del comportamiento “típico” del modelo, es necesario realizar varias repeticiones o réplicas de la simulación. En cada una de estas réplicas se usará, para cada variable aleatoria, una secuencia de observaciones independiente de las secuencias usadas en las restantes réplicas.

A continuación se muestran los resultados obtenidos en una segunda réplica de la simulación. Se han empleado los mismos valores para la condición de finalización y la media de las distribuciones de probabilidad, pero diferentes secuencias de números pseudo aleatorios. La ejecución de la función *randomize()*, en la rutina de inicialización, permite obtener secuencias independientes de números pseudo aleatorios con cada nueva ejecución de la simulación.

Entrada de datos:

```
Media del intervalo entre llegadas:          1.0
Media del tiempo de servicio:              0.5
Num. clientes cuyo tiempo de espera debe  simularse (1-65535):  5
```

A continuación se muestra el obtenido de esta segunda réplica de la simulación. La evolución del sistema está representada en la Figura 5.3.

```
----- Reloj: 0.000000
Empleado LIBRE 0 clientes en la cola Ultimo evento: 0.000000
Cola: Cola vacia
Lista de eventos: Llegada = 4.549548 Marcha = -
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 0
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.549548
Empleado OCUPADO 0 clientes en la cola Ultimo evento: 4.549548
Cola: Cola vacia
Lista de eventos: Llegada = 4.731661 Marcha = 4.584194
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 1
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.584194
Empleado LIBRE 0 clientes en la cola Ultimo evento: 4.584194
Cola: Cola vacia
Lista de eventos: Llegada = 4.731661 Marcha = -
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 1
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.731661
Empleado OCUPADO 0 clientes en la cola Ultimo evento: 4.731661
Cola: Cola vacia
Lista de eventos: Llegada = 4.785454 Marcha = 6.048357
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 4.785454
Empleado OCUPADO 1 clientes en la cola Ultimo evento: 4.785454
Cola: 4.785454
Lista de eventos: Llegada = 5.448385 Marcha = 6.048357
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.000000

----- Reloj: 5.448385
Empleado OCUPADO 2 clientes en la cola Ultimo evento: 5.448385
Cola: 4.785454 5.448385
Lista de eventos: Llegada = 5.829074 Marcha = 6.048357
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 0.662931

----- Reloj: 5.829074
Empleado OCUPADO 3 clientes en la cola Ultimo evento: 5.829074
Cola: 4.785454 5.448385 5.829074
Lista de eventos: Llegada = 5.994754 Marcha = 6.048357
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 1.424310

----- Reloj: 5.994754
Empleado OCUPADO 4 clientes en la cola Ultimo evento: 5.994754
Cola: 4.785454 5.448385 5.829074 5.994754
Lista de eventos: Llegada = 6.400387 Marcha = 6.048357
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 2
  Tiempo total de espera de los clientes en la cola: 0.000000
  Integral num. clientes en la cola respecto del tiempo: 1.921350
```

```

----- Reloj: 6.048357
Empleado OCUPADO 3 clientes en la cola Ultimo evento: 6.048357
Cola: 5.448385 5.829074 5.994754
Lista de eventos: Llegada = 6.400387 Marcha = 7.292892
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 3
  Tiempo total de espera de los clientes en la cola: 1.262903
  Integral num. clientes en la cola respecto del tiempo: 2.135758

----- Reloj: 6.400387
Empleado OCUPADO 4 clientes en la cola Ultimo evento: 6.400387
Cola: 5.448385 5.829074 5.994754 6.400387
Lista de eventos: Llegada = 8.212457 Marcha = 7.292892
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 3
  Tiempo total de espera de los clientes en la cola: 1.262903
  Integral num. clientes en la cola respecto del tiempo: 3.191851

----- Reloj: 7.292892
Empleado OCUPADO 3 clientes en la cola Ultimo evento: 7.292892
Cola: 5.829074 5.994754 6.400387
Lista de eventos: Llegada = 8.212457 Marcha = 7.492419
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 4
  Tiempo total de espera de los clientes en la cola: 3.107409
  Integral num. clientes en la cola respecto del tiempo: 6.761869

----- Reloj: 7.492419
Empleado OCUPADO 2 clientes en la cola Ultimo evento: 7.492419
Cola: 5.994754 6.400387
Lista de eventos: Llegada = 8.212457 Marcha = 7.887635
Cont. estadisticos:
  Numero de clientes que han empezado a ser atendidos: 5
  Tiempo total de espera de los clientes en la cola: 4.770754
  Integral num. clientes en la cola respecto del tiempo: 7.360450

```

Fin de la simulacion...

I N F O R M E

```

-----
Tiempo medio de espera en la cola : 0.954151
Numero medio de clientes en la cola: 0.982386

```

El tiempo medio de espera en la cola se calcula promediando el tiempo de espera de los 3 clientes que han esperado en la cola y la han abandonado antes de que finalice la simulación:

| Entrada en la cola | Salida de la cola | Tiempo de espera |
|--|--------------------------|-------------------------|
| 4.785454 min | 6.048357 min | 1.262903 min |
| 5.448385 min | 7.292892 min | 1.844507 min |
| 5.829074 min | 7.492419 min | 1.663345 min |
| Tiempo total de espera, $D(5)$: | | 4.770755 min |
| Tiempo medio de espera, $\hat{d}(5)$: | | 0.954151 min/cliente |

El valor del área bajo la curva $Q(t)$, es decir, $R(t)$, se ha calculado de la forma siguiente:

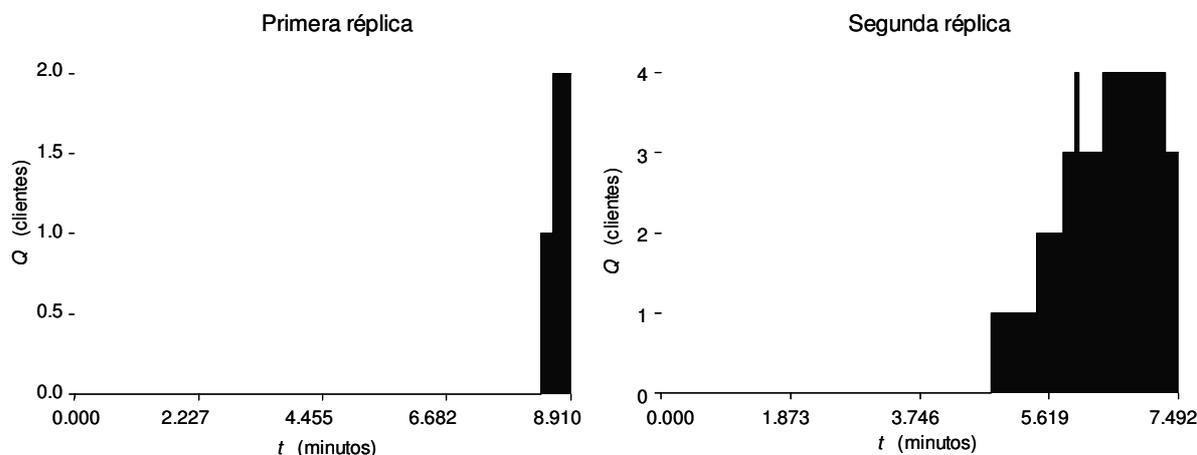


Figura 5.4: $Q(t)$ vs t para cada una de las dos réplicas.

| Intervalo de tiempo (min) | Longitud del intervalo (min) | Clientes en la cola (clientes) | Contribución a la integral (clientes*min) |
|---------------------------|------------------------------|--------------------------------|---|
| 0.000000 – 4.785454 | 4.785454 | 0 | 0.000000 |
| 4.785454 – 5.448385 | 0.662931 | 1 | 0.662931 |
| 5.448385 – 5.829074 | 0.380689 | 2 | 0.761378 |
| 5.829074 – 5.994754 | 0.165680 | 3 | 0.497040 |
| 5.994754 – 6.048357 | 0.053603 | 4 | 0.214412 |
| 6.048357 – 6.400387 | 0.352030 | 3 | 1.056090 |
| 6.400387 – 7.292892 | 0.892505 | 4 | 3.570020 |
| 7.292892 – 7.492419 | 0.199527 | 3 | 0.598581 |
| | | $R(7.492419):$ | 7.360352 |
| | | $\hat{q}(7.492419):$ | 0.982386 |

Donde $R(7.492419)$ representa el valor de la integral y $\hat{q}(7.492419)$ el número medio de clientes en la cola.

Hay algunos estudios de simulación en los cuales el objetivo es investigar el comportamiento del modelo en el estado estacionario, esto es, una vez transcurrido un tiempo simulado muy largo (teóricamente infinito). Supongamos que ésta fuera la finalidad de nuestro estudio. En este caso, es posible calcular analíticamente el tiempo medio de espera en la cola y el número medio de clientes en la cola, obteniéndose, para nuestros valores de las medias de las distribuciones de entrada, los siguientes resultados:

$$\begin{aligned} d(\infty) &= 0.5 \text{ minutos/cliente} \\ q(\infty) &= 0.5 \text{ clientes} \end{aligned} \tag{5.6}$$

Puede observarse que los resultados obtenidos en cualquiera de las dos réplicas anteriores de la simulación distan mucho de los verdaderos en el estacionario. Esto es debido a que se ha realizado la simulación para un número demasiado pequeño de clientes. Asimismo, los resultados obtenidos en ambas réplicas difieren considerablemente entre sí:

| | Primera réplica | Segunda réplica |
|---------------------|------------------------------|---|
| $\hat{d}(5)$ | $= 0.105851$ minutos/cliente | $\hat{d}(5) = 0.954151$ minutos/cliente |
| $\hat{q}(8.597755)$ | $= 0.094473$ clientes | $\hat{q}(7.492419) = 0.982386$ clientes |

La Figura 5.4 muestra la evolución temporal, en cada una de las dos réplicas, del número de clientes de la cola, $Q(t)$.

A continuación se realizan 10 nuevas réplicas de la simulación. Se toman como condiciones iniciales las mismas que en las dos réplicas anteriores: la media del intervalo entre llegadas igual a 1 minuto y la media del tiempo de servicio igual a 0.5 minutos. Sin embargo, en este caso la condición de finalización es que el número de clientes que comienzan a ser atendidos sea $n = 1000$. Se obtienen los resultados siguientes:

| Núm. réplica | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Tiempo medio de espera (min/cliente) | 0.570 | 0.490 | 0.441 | 0.467 | 0.559 | 0.523 | 0.335 | 0.493 | 0.588 | 0.585 |
| Tamaño medio de la cola (clientes) | 0.604 | 0.475 | 0.416 | 0.466 | 0.570 | 0.527 | 0.326 | 0.477 | 0.601 | 0.627 |

La media de los resultados de estas diez repeticiones de la simulación es:

Tiempo medio de espera: 0.505 minutos/cliente

Tamaño medio de la cola: 0.509 clientes

En consecuencia, si el objetivo del estudio de simulación fuera estimar estas medidas en el estacionario, la elección $n = 1000$ como condición de finalización proporcionaría unos valores relativamente próximos a los verdaderos. Sin embargo, no debemos olvidar que la condición de finalización ha sido impuesta arbitrariamente. En la práctica, debe realizarse un estudio para determinar qué condición de finalización proporciona una estimación adecuada de las características en el estacionario. Este punto se discutirá en los próximos temas.

Abundando en este tema, supóngase que se repite la simulación para una frecuencia de llegada de clientes un 80% mayor, es decir:

Entrada de datos:

Media del intervalo entre llegadas: 0.556 minutos
 Media del tiempo de servicio: 0.500 minutos
 Num. de clientes cuyo tiempo de espera desea simularse (1-65535): 1000

El resultado exacto en el estacionario (calculado analíticamente) para estas distribuciones de probabilidad de entrada es:

$$\begin{aligned} d(\infty) &= 4.5 \text{ minutos/cliente} \\ q(\infty) &= 8.1 \text{ clientes} \end{aligned} \quad (5.7)$$

Realizando diez réplicas de la simulación para estas nuevas condiciones de las distribuciones de entrada, se obtienen estos resultados:

| Núm. réplica | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------------------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| Tiempo medio de espera (min/cliente) | 5.692 | 2.664 | 3.739 | 2.530 | 4.349 | 4.626 | 2.844 | 4.059 | 3.666 | 6.954 |
| Tamaño medio de la cola (clientes) | 10.434 | 4.886 | 6.624 | 4.640 | 7.403 | 8.824 | 5.175 | 7.521 | 6.582 | 12.831 |

La media de los resultados de estas diez nuevas repeticiones de la simulación es:

Tiempo medio de espera: 4.112 minutos/cliente

Tamaño medio de la cola: 7.492 clientes

Los cuales son sensiblemente diferentes de los verdaderos. En consecuencia, parece claro que la condición de finalización no puede escogerse arbitrariamente.

En este ejemplo, la simulación finaliza cuando el número de clientes que han comenzado a ser atendidos supera un valor preestablecido ($n = 5$, $n = 1000$, etc.) con lo cual el valor del reloj de la simulación en el instante final de ésta es una variable aleatoria. Son posibles otros muchos tipos de condición de finalización. Por ejemplo:

- Que el reloj de la simulación alcance un valor determinado, por ejemplo, 480 minutos. Es decir, las 8 horas que dura la jornada laboral del empleado. En este caso, el número de clientes que han comenzado a ser atendidos durante la simulación, n , es una variable aleatoria.
- Que el reloj de la simulación alcance un valor igual o superior a 480 minutos y además el número de clientes en la cola sea cero y el empleado esté desocupado. Esta situación correspondería con la siguiente "estrategia de cierre" de la oficina: finalizada la jornada laboral de 8 horas, el empleado impide la entrada de nuevos clientes, pero prosigue con su trabajo hasta que atiende a todos los clientes que en ese instante se encuentran dentro de la oficina. Con esta condición de finalización, tanto el valor del reloj de la simulación en el instante final de la simulación como el número de clientes que han comenzado a ser atendidos hasta ese instante son variables aleatorias.

5.4. Elección de la política de gestión del inventario de un único producto

En esta sección se analiza el funcionamiento de una compañía dedicada a la comercialización de cierto producto. La compañía actúa como intermediaria entre el proveedor y los clientes. A fin de poder satisfacer rápidamente las peticiones de los clientes, la compañía dispone de un almacén para el producto. La clave de la rentabilidad de la compañía está en realizar una adecuada gestión del inventario de este almacén. Debe intentar satisfacer las demandas de los clientes, pero manteniendo el nivel de inventario lo más bajo posible, con el fin de minimizar los costes de almacenamiento.

La estructura lógica del modelo

La política de atención al cliente es la siguiente. Los clientes solicitan los pedidos atendiendo únicamente a sus necesidades: los intervalos de tiempo entre pedidos consecutivos son variables y el número de unidades de que se compone el pedido también. En cierto instante de tiempo, t , la compañía tiene en el almacén cierto número de unidades del producto. Este número se representa: $I(t)$. Si un cliente solicita a la compañía una cantidad de producto no superior a la almacenada, la demanda es atendida en el momento. Por el contrario, si la cantidad demandada supera a la que la compañía tiene almacenada en ese instante, se sirve inmediatamente la totalidad de lo almacenado, satisfaciéndose el exceso de demanda cuando la llegada de nuevo producto lo permita. En este caso, la nueva cantidad de producto inventariada (la que había en el almacén menos la demandada) será negativa. Cuando la compañía recibe la mercancía del proveedor, se satisfacen las demandas pendientes y el producto restante se almacena.

Al principio de cada mes, la compañía hace una evaluación de su inventario y decide qué cantidad de producto, Z , debe solicitar al proveedor. El valor de Z se obtiene de aplicar la fórmula siguiente:

$$Z = \begin{cases} S - I & \text{si } I < s \\ 0 & \text{si } I \geq s \end{cases} \quad (5.8)$$

El parámetro I es el número de unidades del producto que tiene almacenadas la compañía en el momento de ordenar el pedido. Los parámetros s y S son los que establecen la política de gestión del inventario de la empresa. Esta política tiene una repercusión directa sobre los resultados económicos de la compañía, con lo cual la elección de estos parámetros debe ser objeto de un cuidadoso estudio. Para ello, deben tenerse en cuenta que existen tres contribuciones al coste económico de la actividad de la compañía:

- Los *costes del pedido*, calculados como $C = K + i \cdot Z$. El precio de cada unidad de producto son i euros. Existen unos gastos fijos de tramitación, independientes del número de unidades del pedido, que debe satisfacer la compañía: K euros. El tiempo que transcurre desde que la compañía solicita un pedido y lo recibe es variable.
- Los *costes de almacenamiento*. El coste medio mensual (euros/mes) que conlleva el almacenamiento del inventario es:

$$h \cdot \frac{\int_0^n I^+(t) \cdot dt}{n} \quad (5.9)$$

donde $I^+(t) = \max\{I(t), 0\}$. El parámetro h es el coste de almacenamiento por unidad de producto y por mes. El valor $\frac{\int_0^n \max\{I(t), 0\} \cdot dt}{n}$ es el promedio temporal del número de unidades de producto en inventario. Este promedio se calcula considerando la evolución del inventario, $I(t)$, a lo largo de n meses (la unidad en que se mide el tiempo para el cálculo de la integral es el mes). Se ha tenido en cuenta que el inventario puede tomar valores negativos. Mientras el inventario sea negativo, a efectos de cálculo del coste de almacenamiento, se considera que vale cero.

- Los *costes del déficit de inventario*. Este coste pretende medir la pérdida de prestigio frente al cliente que supone no poder servirle inmediatamente la totalidad de su pedido. Este coste se calcula de la forma siguiente:

$$p \cdot \frac{\int_0^n I^-(t) \cdot dt}{n} \quad (5.10)$$

donde $I^-(t) = \max\{-I(t), 0\}$. El parámetro p representa el coste por unidad de producto no servido y por mes. El valor $\frac{\int_0^n \max\{-I(t), 0\} \cdot dt}{n}$ es el promedio temporal del número de unidades de producto demandadas por los clientes y que no han podido ser servidas por falta de inventario.

Objetivo del estudio de simulación

El objetivo del estudio de simulación es encontrar qué valores de los parámetros s y S hacen que la operación de la compañía sea más rentable. Para ello, se simula el inventario durante un periodo de n meses y se compara el coste total medio mensual de cada una de las diferentes políticas de gestión del inventario. El coste total es la suma de los costes de los pedidos, de almacenamiento y el coste de no poder satisfacer la demanda. Asimismo, se desea conocer la contribución al coste total de cada uno de los tres factores: pedidos, almacenamiento y demanda insatisfecha. Se pretende investigar los siguientes valores de los parámetros de la gestión del inventario:

| | | | | | | | | | |
|-----|----|----|----|-----|----|----|-----|----|-----|
| s | 20 | 20 | 20 | 20 | 40 | 40 | 40 | 60 | 60 |
| S | 40 | 60 | 80 | 100 | 60 | 80 | 100 | 80 | 100 |

Para cada una de estas políticas de gestión del inventario, el objetivo de la simulación es estimar las cuatro variables siguientes:

- El coste medio total mensual.
- El coste medio mensual de los pedidos.
- El coste medio mensual del almacenamiento.
- El coste medio mensual de la demanda insatisfecha.

Aspectos matemáticos del modelo

Un estudio acerca de las actividades de la compañía permite realizar las siguientes hipótesis de modelado:

- Los intervalos de tiempo entre pedidos sucesivos son variables aleatorias IID (independientes e idénticamente distribuidas) exponenciales de media 0.1 meses. Estas variables aleatorias, R , pueden generarse a partir de una variable aleatoria uniforme de la forma siguiente:

$$\left. \begin{array}{l} X \sim U(0, 1) \\ r = -\beta \cdot \ln(x) \end{array} \right| \rightarrow R \sim \text{expo}(\beta) \quad (5.11)$$

- El número de unidades de que constan los pedidos que los clientes hacen a la compañía, D , son variables aleatorias IID, independientes de los instantes en que se producen, con la siguiente probabilidad:
 - $d = 1$ tiene probabilidad $\frac{1}{6}$
 - $d = 2$ tiene probabilidad $\frac{1}{6}$
 - $d = 3$ tiene probabilidad $\frac{1}{6}$
 - $d = 4$ tiene probabilidad $\frac{1}{6}$

Las observaciones de estas variables aleatorias, D , pueden generarse a partir de observaciones de una variable aleatoria uniforme, $X \sim U(0, 1)$, de la forma siguiente. Si la observación de X toma un valor en el intervalo $[0, \frac{1}{6})$, se asigna a la observación de D el valor 1: $d = 1$. Si la observación de X toma un valor en el intervalo $[\frac{1}{6}, \frac{2}{6})$, se asigna $d = 2$, y así sucesivamente.

- El tiempo de servicio del proveedor, es decir, el tiempo que transcurre desde que la compañía hace el pedido hasta que recibe el producto, es una variable aleatoria, V , uniformemente distribuida en el intervalo 0.5 meses, 1 mes. Pueden generarse observaciones de V de la forma siguiente:

$$\left. \begin{array}{l} X \sim U(0, 1) \\ v = \frac{1}{2} + (1 - \frac{1}{2}) \cdot x \end{array} \right| \rightarrow V \sim U\left(\frac{1}{2}, 1\right) \quad (5.12)$$

Asimismo, el modelo consta de los siguientes parámetros:

- Los parámetros s y S de gestión del inventario.
- Los coeficientes h y p de las expresiones del coste de almacenamiento y de demanda insatisfecha. Sus valores son: $h = 100 \frac{\text{euros}}{\text{unidad de producto} \times \text{mes}}$ y $p = 500 \frac{\text{euros}}{\text{unidad de producto} \times \text{mes}}$.
- Precios que impone el proveedor. Estos son: tramitación del pedido $K = 3200$ euros y precio por unidad $i = 300$ euros.

Definición del experimento

Las *condiciones iniciales* del experimento son las siguientes:

- Existen 60 unidades del producto en inventario, $I(0) = 60$.
- No hay ningún pedido previsto.

La *condición de finalización* de la simulación es que el reloj alcance 120 meses, es decir, $n = 120$.

Modelado del funcionamiento del sistema

Son necesarias las siguientes tres variables para describir el estado del sistema:

- El número de unidades de producto almacenadas, $I(t)$.
- El número de unidades de producto que la compañía ha encargado a su proveedor y que está a la espera de recibir, Z .
- El instante de tiempo en que se produjo el último cambio en el inventario, t_{evento} . Esta variable debe conocerse a fin de actualizar el cálculo de las integrales $\int_0^n I^+(t) \cdot dt$ y $\int_0^n I^-(t) \cdot dt$.

El modelo contiene cuatro tipos de eventos (ver Figuras 5.5 y 5.6):

- **Llegada a la compañía del producto solicitado al proveedor.** Las acciones asociadas a este evento son las siguientes:
 - Actualización del cálculo de las integrales $\int_0^n I^+(t) \cdot dt$ y $\int_0^n I^-(t) \cdot dt$.
 - Incrementar el inventario, $I(t)$, de acuerdo a la cantidad de producto recibida, Z .
 - Asignar el valor NO_DEFINIDO, en la lista de eventos, al instante en que se producirá la próxima recepción de producto.
- **Un cliente realiza un pedido.** Las acciones asociadas a este evento son las siguientes:
 - Este evento implica una disminución del inventario, por consiguiente debe actualizarse el cálculo de las integrales $\int_0^n I^+(t) \cdot dt$ y $\int_0^n I^-(t) \cdot dt$.
 - Generación de un valor aleatorio para la cuantía del pedido.
 - Actualizar el inventario, restando el número de unidades de que consta el pedido.
 - Generación de un valor aleatorio para el intervalo de tiempo entre los pedidos de los clientes. El instante en que se producirá el siguiente evento de este tipo de calcula sumando el valor generado al del reloj de la simulación.
- **La compañía evalúa su inventario.** Al comienzo de cada mes la compañía decide si debe o no encargar la compra de producto y en qué cantidad. La decisión depende del inventario en ese instante y del valor de los parámetros s y S . Las acciones asociadas al evento son las siguientes:
 - Si el número de unidades almacenadas de producto es inferior a s , se realiza un pedido al proveedor de $S - I(t)$ unidades. Debe actualizarse el valor de los costes de pedido. Además, debe generarse el valor del intervalo aleatorio de tiempo entre que se solicita el pedido hasta que se recibe. Sumado el valor generado al del reloj de la simulación se obtiene el instante de recepción del pedido.
 - Si el número de unidades almacenadas de producto es mayor o igual que s , ese mes no se realiza pedido.
- **Final de la simulación.** Este evento se activa cuando el reloj de la simulación toma el valor n meses. Las acciones asociadas a este evento son las siguientes:
 - Actualización del cálculo de las integrales $\int_0^n I^+(t) \cdot dt$ y $\int_0^n I^-(t) \cdot dt$.
 - Generación de los informes.

La *rutina de tiempo* del programa de simulación gestiona el avance del reloj de la simulación y la activación de los eventos. Asimismo, debe establecer, en aquellos casos en que sea relevante, en qué orden deben ejecutarse los eventos si se activan varios en el mismo instante de tiempo. Por ejemplo, si coinciden el instante de evaluación del inventario con el final de la simulación, se dará prioridad al evento de finalización de la simulación.

A fin de calcular las variables de salida, es preciso definir los contadores estadísticos siguientes:

- Coste total de los pedidos realizados por la compañía hasta ese momento. Este contador es necesario a fin de calcular el coste medio mensual debido a los pedidos.
- Área bajo la función $I^+(t)$. Este estadístico es necesario para calcular el coste medio mensual de almacenamiento.
- Área bajo la función $I^-(t)$. Es necesario para calcular el coste medio mensual de la demanda insatisfecha.

En el Anexo 2, situado al final del Tema, se muestra el código fuente del programa de simulación. A continuación se muestra un ejemplo de la evolución del sistema durante una

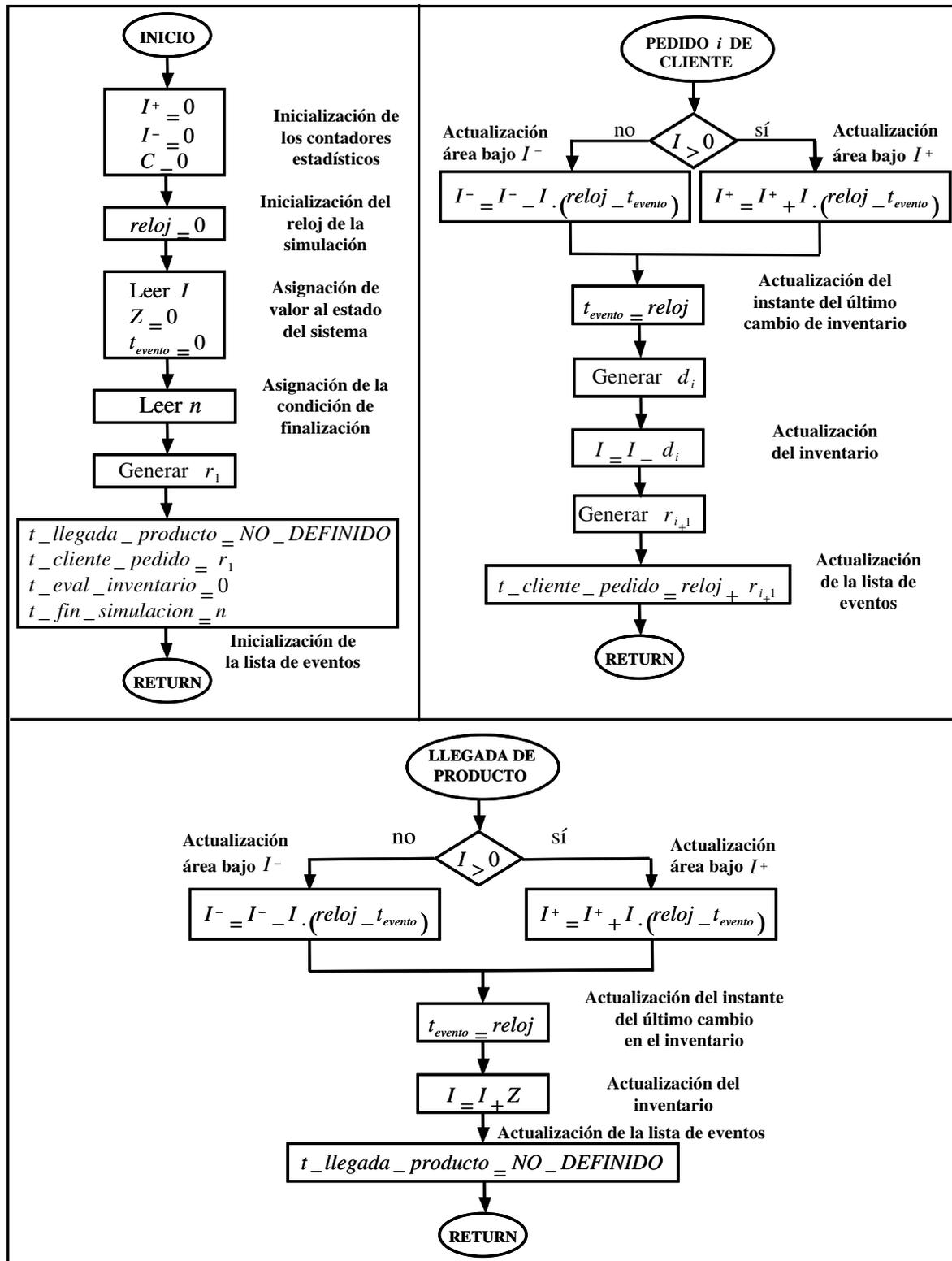


Figura 5.5: Flujo de acciones de las rutinas de inicialización y de eventos.

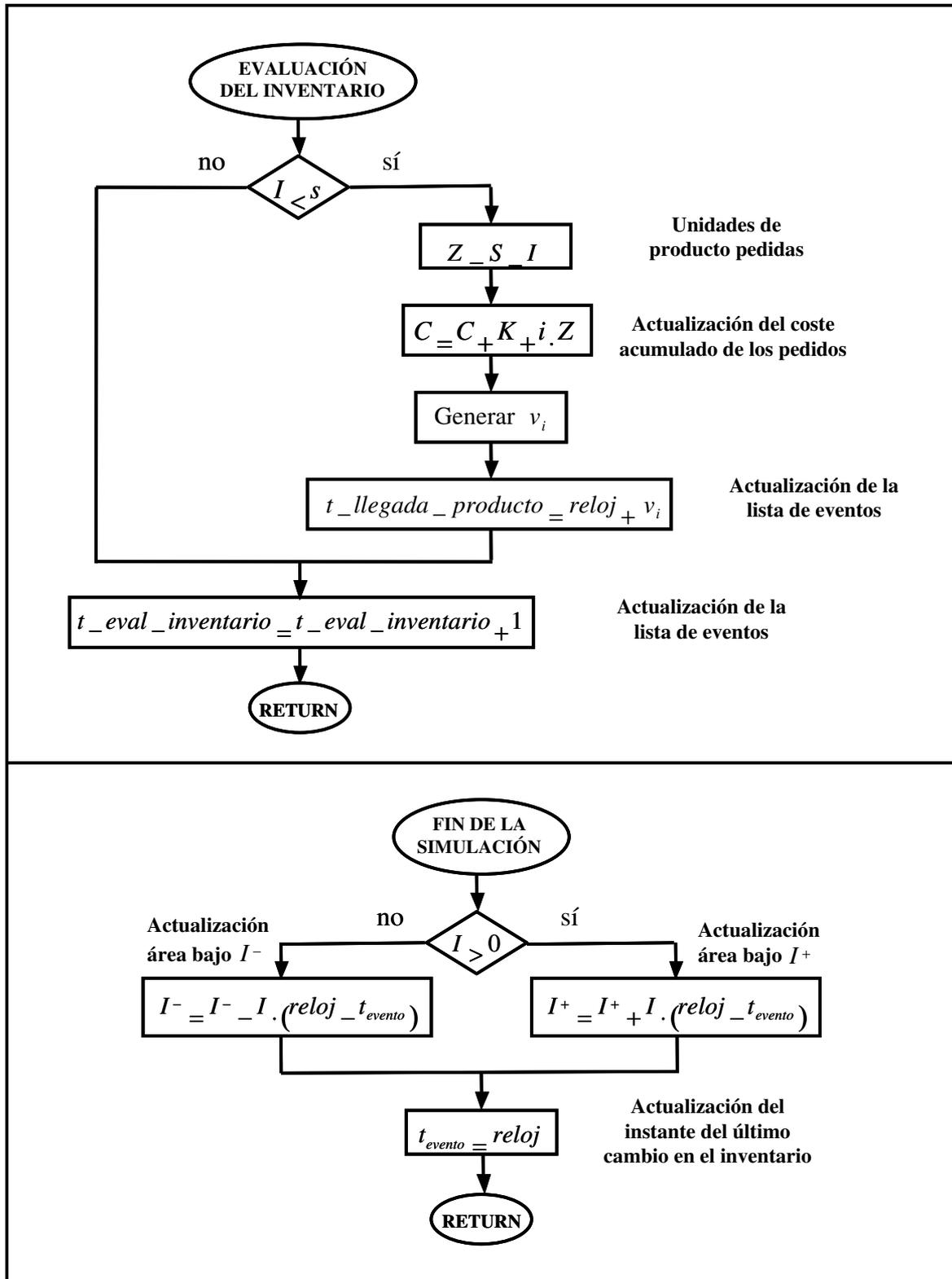


Figura 5.6: Flujo de acciones de las rutinas de eventos.

réplica de la simulación y el resultado obtenido. Los valores asignados a las variables de entrada son los siguientes:

```
Entrada de datos:
-----
Parametros de gestion del inventario:
    Parametro s: 20
    Parametro S: 40
Duracion de la simulacion (meses): 6
```

La condición de la finalización de la simulación es que el reloj alcance el valor 6 meses. El tiempo se mide en meses. En la Figura 5.7 se muestra la evolución del inventario y del producto que la compañía encarga a su proveedor, así como de los tres contadores estadísticos: el coste de los pedidos que la compañía realiza y las integrales de I^+ y de I^- .

El estudio de simulación

Como se ha indicado anteriormente, el objetivo del estudio de simulación es comparar las nueve siguientes políticas de gestión del inventario:

| | | | | | | | | | |
|-----|----|----|----|-----|----|----|-----|----|-----|
| s | 20 | 20 | 20 | 20 | 40 | 40 | 40 | 60 | 60 |
| S | 40 | 60 | 80 | 100 | 60 | 80 | 100 | 80 | 100 |

Para ello, se van a realizar 5 réplicas de cada una de estas condiciones experimentales. En todos los casos, la condición de finalización es que el reloj de la simulación alcance 120 meses. Obsérvese la arbitrariedad en la elección tanto del número de réplicas de la simulación de cada política, como en la elección de la condición de finalización. En la Tabla 5.4 se muestran los resultados obtenidos. La última columna de cada tabla contiene el valor medio de las 5 réplicas. La terminología empleada es la siguiente:

- *Simulación*: Número de réplica.
- *Pedidos*: Coste medio mensual de los pedidos al suministrador.
- *Manten.*: Coste medio mensual del mantenimiento del inventario.
- *Déficit*: Coste medio mensual de la demanda insatisfecha.
- *Total*: Coste medio mensual total.
- *Media*: Valor medio de las 5 réplicas.

Atendiendo al valor medio del coste total obtenido en las 5 réplicas anteriores, parece que la política de inventario $s = 20$, $S = 60$ es preferible a las demás estudiadas. Sin embargo, cabe preguntarse si comparar los valores medios del coste total mensual es el procedimiento de análisis más adecuado ¿En qué medida son significativas las diferencias entre una y otra política?. Asimismo, ¿hasta qué punto son fiables los resultados obtenidos? ¿Cómo influye el número de réplicas de la simulación, que arbitrariamente hemos fijado en 5, en el resultado? ¿Y la condición de finalización de la simulación, que arbitrariamente hemos establecido en 120 meses?. En consecuencia, hasta que no respondamos a estas preguntas no podremos concluir que una determinada política es mejor (con cierto tanto por ciento de confianza) que las demás.

Los datos de salida de la simulación son comúnmente analizados empleando métodos de *análisis gráfico* y métodos de *análisis numérico*. El análisis gráfico suele realizarse en primer lugar: resulta muy eficaz, intuitivo y sencillo, y permite identificar propiedades interesantes que pueden orientar los análisis siguientes. De hecho, los analistas de datos experimentados invierten la mayoría del tiempo haciendo análisis gráfico. El primer paso del análisis gráfico consiste en la representación de los datos de salida “en bruto”.

Cuando el objetivo del estudio de simulación es la comparación de diferentes alternativas, una herramienta de análisis gráfico de probada eficacia son los *boxplots*. Los datos obtenidos

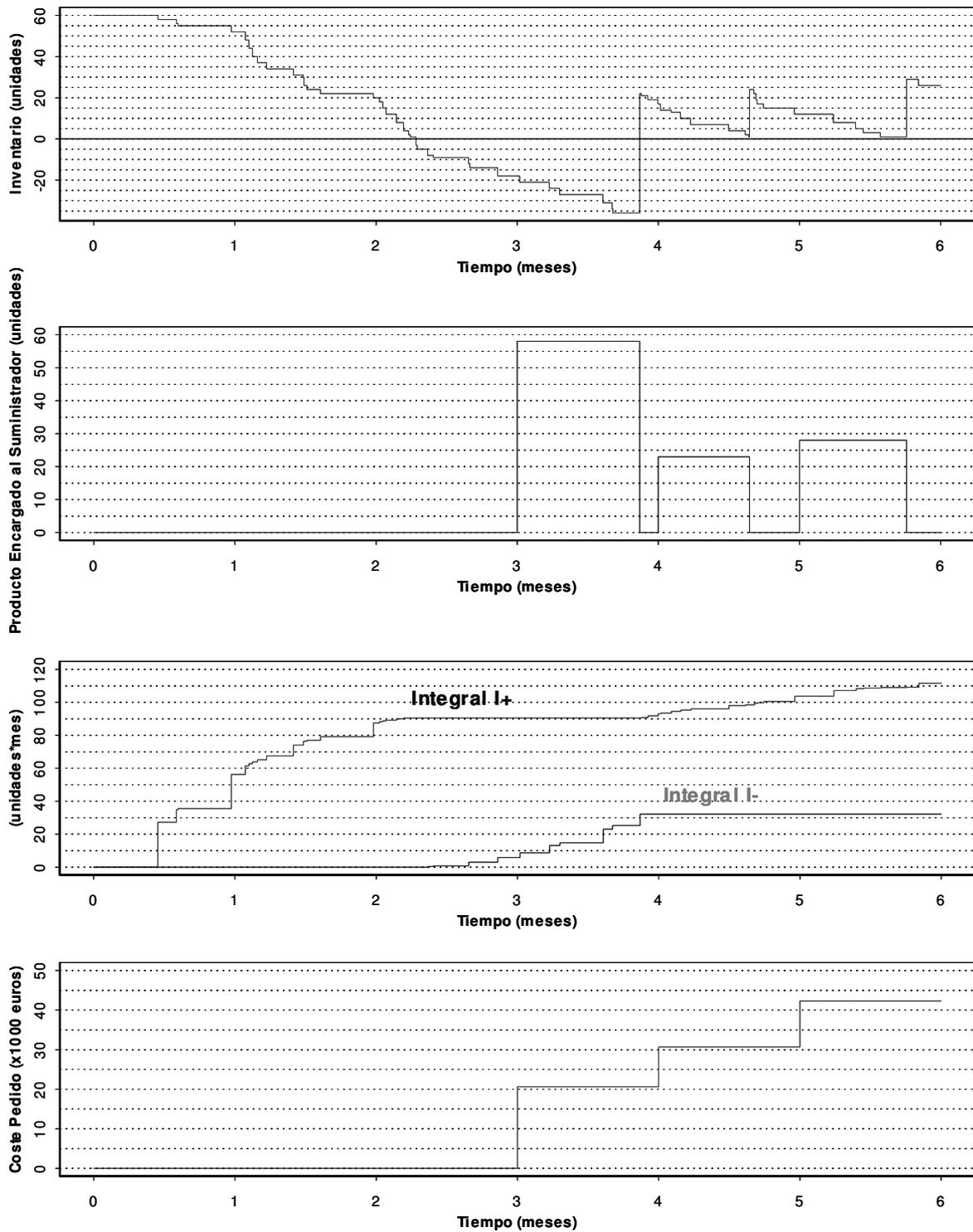


Figura 5.7: Evolución de algunas variables del modelo durante la simulación.

Tabla 5.2: Resultados obtenidos en 5 réplicas de cada condición experimental.

| | | | | | | | |
|-------------------|---------|---------|---------|---------|---------|---------|---------|
| $s = 20, S = 40$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 9952.5 | 10345.8 | 10174.2 | 9876.7 | 9.755.8 | 10021.0 |
| | Manten. | 910.2 | 841.1 | 845.6 | 888.0 | 857.4 | 868.5 |
| | Déficit | 1938.0 | 2064.2 | 2180.8 | 2184.8 | 1944.5 | 2062.5 |
| | Total | 12810.7 | 13251.1 | 13200.6 | 12949.4 | 12557.6 | 12953.8 |
| $s = 20, S = 60$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 8885.0 | 8797.5 | 9022.5 | 8712.5 | 8707.5 | 8825.0 |
| | Manten. | 1680.0 | 1788.4 | 1711.8 | 1761.3 | 1920.9 | 1772.5 |
| | Déficit | 1444.5 | 1366.3 | 1562.5 | 1090.2 | 1254.1 | 1343.5 |
| | Total | 12009.5 | 11952.2 | 12296.8 | 11564.1 | 11882.5 | 11941.0 |
| $s = 20, S = 80$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 7946.7 | 8604.2 | 8536.7 | 8606.7 | 8020.0 | 8342.9 |
| | Manten. | 2855.8 | 2628.9 | 2709.6 | 2541.4 | 2867.4 | 2720.6 |
| | Déficit | 896.4 | 1116.0 | 1080.3 | 1133.9 | 653.4 | 976.0 |
| | Total | 11698.9 | 12349.1 | 12326.5 | 12282.0 | 11540.9 | 12039.5 |
| $s = 20, S = 100$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 8184.2 | 8144.2 | 8646.7 | 8116.7 | 8909.2 | 8400.2 |
| | Manten. | 3700.4 | 3602.7 | 3621.1 | 3778.1 | 3642.3 | 3668.9 |
| | Déficit | 647.8 | 815.3 | 666.9 | 592.4 | 836.0 | 711.7 |
| | Total | 12532.4 | 12562.2 | 12934.6 | 12487.2 | 13387.4 | 12780.8 |
| $s = 40, S = 60$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 9886.7 | 10053.3 | 10120.0 | 9791.7 | 9619.2 | 9894.2 |
| | Manten. | 2392.2 | 2556.0 | 2523.1 | 2574.6 | 2574.7 | 2524.1 |
| | Déficit | 367.6 | 201.6 | 197.8 | 120.9 | 272.8 | 232.1 |
| | Total | 12642.5 | 12811.0 | 12840.9 | 12487.1 | 12466.7 | 12649.6 |
| $s = 40, S = 80$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 8432.5 | 8990.0 | 8850.0 | 9025.0 | 8950.0 | 8849.5 |
| | Manten. | 3556.3 | 3466.8 | 3508.6 | 3505.4 | 3422.4 | 3491.9 |
| | Déficit | 136.6 | 225.6 | 145.6 | 124.8 | 188.0 | 164.1 |
| | Total | 12125.4 | 12682.3 | 12504.1 | 12655.3 | 12560.5 | 12505.5 |
| $s = 40, S = 100$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 8576.7 | 8515.0 | 8578.3 | 9139.2 | 8155.0 | 8592.8 |
| | Manten. | 4351.8 | 4488.1 | 4498.0 | 4314.5 | 4523.5 | 4435.2 |
| | Déficit | 167.1 | 141.3 | 143.8 | 78.3 | 58.0 | 117.7 |
| | Total | 13095.6 | 13144.4 | 13220.1 | 13532.0 | 12736.6 | 13145.5 |
| $s = 60, S = 80$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 10691.7 | 10075.8 | 9952.5 | 9621.7 | 9793.3 | 10027.0 |
| | Manten. | 4316.9 | 4462.7 | 4498.0 | 4495.6 | 4545.6 | 4463.8 |
| | Déficit | 6.9 | 18.4 | 11.3 | 5.3 | 0.0 | 8.4 |
| | Total | 15015.5 | 14556.8 | 14461.8 | 14122.5 | 14338.9 | 14499.1 |
| $s = 60, S = 100$ | Réplica | 1 | 2 | 3 | 4 | 5 | Media |
| | Pedidos | 8575.8 | 9135.8 | 8946.7 | 8624.2 | 8842.5 | 8825.0 |
| | Manten. | 5658.5 | 5436.5 | 5485.8 | 5646.7 | 5631.8 | 5571.9 |
| | Déficit | 0.0 | 5.3 | 0.1 | 1.0 | 0.5 | 1.4 |
| | Total | 14234.3 | 14577.6 | 14432.6 | 14271.9 | 14474.8 | 14398.2 |

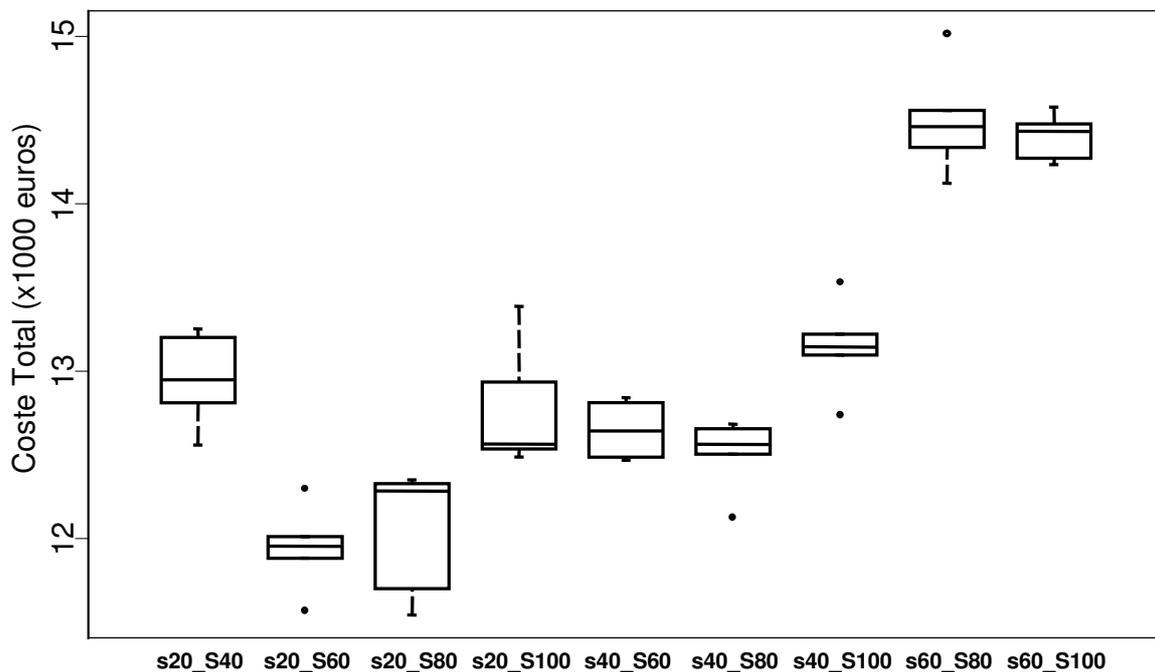


Figura 5.8: Representación mediante boxplots de los resultados del estudio.

de simular cada alternativa se representan mediante un boxplot. La comparación visual entre los boxplots de las diferentes alternativas resulta muy explicativa. Obsérvese, en la Figura 5.8, la representación mediante boxplots de los datos obtenidos para las diferentes políticas de gestión del inventario. ¿Puede extraerse alguna conclusión?

5.5. Modelado mediante grafos de eventos

A medida que se van realizando modelos más complejos, se hace patente la necesidad de disponer de una metodología que sistematice la descripción del modelo. Una de estas metodologías, la cual destaca por su simplicidad, son los grafos de eventos.

Los *grafos de eventos*² son una metodología para representar de manera gráfica los modelos de tiempo discreto.

En esta metodología, los eventos, cada uno representado mediante un *nodo*, son conectados mediante *arcos dirigidos* (flechas), que representan cómo los eventos son planificados a partir de otros eventos o a partir de ellos mismos.

Cada arco dirigido puede tener asociada una condición booleana (denominada *condición del arco*) y/o un retardo en el tiempo. En la Figura 5.9 se muestra la estructura fundamental de los grafos de eventos, que debe ser interpretada de la forma siguiente: una vez ejecutado el evento A (es decir, calculado el estado correspondiente al mismo), si se verifica la condición (i), entonces se planifica el evento B en el calendario de eventos, el cual se disparará transcurrido un tiempo *t*.

²Estas explicaciones están extraídas del artículo (Buss 2001) y del texto (Law & Kelton 2000).

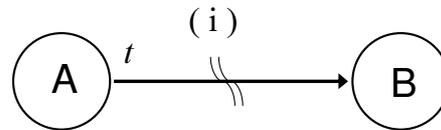


Figura 5.9: Estructura fundamental de los grafos de eventos.

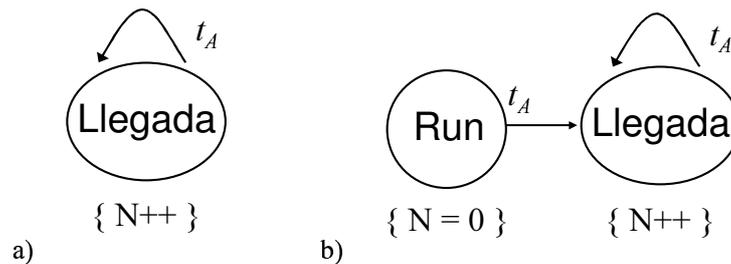


Figura 5.10: Grafo de eventos de un proceso de llegada.

Por convenio, el retardo se escribe en el extremo del arco donde no está la punta de la flecha. La condición del arco se escribe en la parte central del arco, encima de la línea serpenteante que lo corta.

Si en una transición no existe retardo en el tiempo, entonces éste se omite. Por otra parte, si el evento B es planificado siempre que ocurre el evento A, entonces la condición del arco se omite y se dice que el arco es *incondicional*.

Ejemplo 5.5.1. El grafo de eventos más sencillo es un modelo de un proceso de llegada, que consta de un único evento (la llegada) y una única variable de estado: el número total de entidades que han llegado hasta ese momento (N). El intervalo de tiempo entre llegadas sucesivas, t_A , puede ser constante, pueden ser observaciones independientes de variables aleatorias idénticamente distribuidas, etc. Cada vez que se produce la llegada de una entidad, se incrementa la variable de estado: $N++$. El grafo de eventos se muestra en la Figura 5.10a.

Con el fin de inicializar las variables de estado y definir cómo debe arrancar la simulación, se define un evento especial, denominado Run, que es ejecutado en el instante inicial de la simulación. Es decir, este evento se sitúa, en el calendario de eventos, en el instante de tiempo 0.0.

Así pues, para convertir el grafo de eventos mostrado en la Figura 5.10a en un modelo ejecutable, se le añade el evento Run, el cual inicializa la variable de estado y planifica la primera llegada. El grafo se muestra en la Figura 5.10b. \square

Ejemplo 5.5.2. Considérese el modelo de una oficina de información con una única cola, atendida por k empleados. Los clientes llegan de uno en uno. Si en el instante en el que llega el cliente:

- Todos los empleados están ocupados, entonces el cliente espera en una única cola a que uno de los empleados quede libre. La cola es atendida según el orden de llegada. Cuando un empleado queda libre, entonces comienza a atender al primer cliente de la cola.
- Si al menos uno de los empleados está libre, el cliente es atendido.

Una vez a sido atendido, el cliente abandona la oficina.

El intervalo de tiempo entre llegadas sucesivas de los clientes (t_A) y el tiempo de servicio (t_S) son observaciones independientes de sendas variables aleatorias.

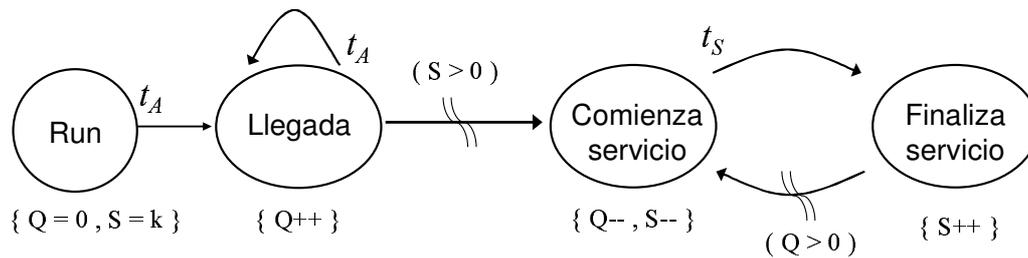


Figura 5.11: Grafo de eventos del sistema descrito en el Ejemplo 5.5.2.

El estado del sistema se define mediante las dos variables siguientes:

- El número de clientes en cola (Q).
- El número de empleados que se encuentran libres (S).

El grafo de eventos se muestra en la Figura 5.11. □

El modelo mostrado en la Figura 5.11 se asemeja a un modelo orientado al proceso, dado que los eventos corresponden con la secuencia de acciones que le suceden a la entidad (el cliente) al progresar por el sistema. Sin embargo, si se observa el modelo de la Figura 5.11 detenidamente, se verá que el arco que va desde el evento *Finaliza servicio* hasta el evento *Comienza servicio* no está presente en el modelo orientado al proceso.

Así pues, es importante tener presente que el grafo de eventos captura la dependencia en las planificaciones de los eventos del modelo, no el flujo de clientes o entidades en el sistema. Es decir, el grafo de eventos no representa un flujo sincrónico de ejecuciones de eventos, sino la relación en la planificación de los eventos, que son disparados de manera asíncrona cuando la simulación es ejecutada.

Eventos simultáneos

Se dice que dos o más eventos son *simultáneos* cuando su disparo está planificado exactamente en el mismo instante de tiempo.

En algunos casos, el orden en el que se ejecutan los eventos simultáneos es irrelevante. Sin embargo, en otros casos, no lo es: cambios en el orden en el que se ejecutan los eventos simultáneos pueden producir resultados de la simulación significativamente diferentes.

Puesto que los ordenadores trabajan con números de precisión finita, esta posibilidad no puede descartarse aun cuando se usen en el modelo variables aleatorias "continuas". Si se usan distribuciones de probabilidad discretas para modelar los retardos en el tiempo, entonces la probabilidad de que se produzcan eventos simultáneos aumenta drásticamente.

La metodología de los grafos de eventos proporciona la capacidad de asignar prioridad a los arcos. Si bien la prioridad de los arcos no suele mostrarse en el grafo, todas los entornos software que soportan la metodología de los grafos de eventos permiten priorizar los arcos.

Simkit

Simkit es un conjunto de paquetes Java que facilitan la programación de modelos descritos empleando la metodología de los grafos de eventos. Se trata de software libre, que puede ser descargado gratuitamente del sitio web:

<http://diana.or.nps.navy.mil/simkit/>

EJERCICIOS

- 5.1** Suponga que en el modelo de la cola atendida por un único empleado no se desea calcular el tiempo medio de espera en la cola. ¿Cómo deberían modificarse los flujos de acciones asociadas a los eventos, mostrados en la Figura 4.2?
- 5.2** En el modelo de la cola atendida por un único empleado desea calcularse una nueva variable de salida: el tiempo de ciclo medio de los clientes. El *tiempo de ciclo* es el tiempo total que pasa el cliente en el sistema, es decir, el tiempo que espera en la cola más el tiempo durante el cual el empleado le atiende. ¿Cómo deberían modificarse las acciones asociadas a los eventos?
- 5.3** Suponga que, además de la nueva variable de salida descrita en el Problema 5.2, se desean calcular:
- El tiempo máximo de espera en la cola.
 - El número máximo de clientes que esperan en la cola.
 - La proporción de clientes que han debido esperar en la cola más de 1 minuto.
 - La “*utilización*” del empleado, que se define como el tiempo que ha permanecido ocupado dividido por el tiempo total (es decir, el tiempo que ha estado ocupado más el tiempo durante el cual ha estado libre).

¿Cómo deberían modificarse las acciones asociadas a los eventos?

- 5.4** Modifique el modelo de la cola atendida por un único de modo que la condición de finalización sea la siguiente. Una vez el sistema ha operado durante 8 horas, ya no se permite el acceso de más clientes. El empleado termina de atender a los clientes que en ese instante se encuentran en el sistema, y una vez ha concluido, finaliza la simulación.
- 5.5** Suponga que en la cola del modelo descrito en el Problema 5.4 sólo hay sitio para que esperen dos clientes. Cuando hay dos clientes en la cola, y llega un nuevo cliente, éste no puede ponerse a la cola, con lo cual abandona directamente el sistema (esta propiedad se denomina *balking*). Modifique el modelo de modo que contemple esta nueva característica, y de modo que se calcule una nueva variable de salida: el número total de clientes que no han podido ser atendidos por estar la cola llena.
- 5.6** Realice la siguiente modificación en el modelo de gestión del inventario. Si en el momento de ordenar el pedido la cantidad de producto almacenada es $I < 0$, entonces la compañía realiza una *orden urgente* a su proveedor. Si la cantidad almacenada es $0 \leq I < s$, entonces se realiza una orden de compra normal.

El coste de tramitación de una orden urgente es superior al de una orden normal, si bien el coste por unidad de producto es el mismo. Una orden urgente de Z unidades de producto cuesta $C_{urgente} = K_{urgente} + i \cdot Z$, donde $K_{urgente} = 7000$ euros. En contrapartida, el tiempo de entrega de una orden urgente está uniformemente distribuido entre 0.25 y 0.5 meses.

- 5.7** En el modelo del inventario, suponga que los productos almacenados son perecederos. Cada unidad de producto tiene una fecha de caducidad, que está distribuida uniformemente entre 1.5 y 2.5 meses, empezando a contar desde que es recibida y se almacena. Obsérvese que cada una de las diferentes unidades de producto pertenecientes a un mismo pedido puede tener una fecha de caducidad diferente. Los productos caducados no tienen ningún valor, y deben ser desechados.

La compañía descubre que una determinada unidad de producto está caducada cuando la examina justo antes de venderla. Los productos del inventario se procesan de manera FIFO, es decir, se envían antes los productos que más tiempo llevan almacenados. Modifique el modelo para describir esta nueva característica, y además añada una nueva variable de salida: el número total de unidades de producto que han debido ser desechadas por encontrarse caducadas.

- 5.8** Realice el modelo de simulación del siguiente sistema. Un servicio de ventas consta de dos empleados dispuestos en serie, cada uno con su propia cola FIFO. El cliente es atendido por el primer empleado, a continuación por el segundo, y seguidamente abandona el sistema. El tiempo que transcurre entre la llegada de un cliente y la del siguiente está distribuido exponencialmente, con media 1 minuto. El tiempo que emplea el

primer empleado en atender a un cliente está distribuido exponencialmente, con media 0.7 minutos; mientras que el tiempo que emplea el segundo empleado está distribuido exponencialmente, con media 0.9 minutos.

Las variables de salida del modelo son:

- El tiempo medio de espera en cada una de las dos colas.
- La longitud media de cada una de las dos colas.
- La *utilización* de cada uno de los empleados.

La condición de finalización es que el tiempo simulado alcance los 1000 minutos. Se supone que la oficina funciona ininterrumpidamente durante todo ese tiempo.

- 5.9** Modifique el modelo del Problema 5.8, de modo que se contemple que el tiempo que tarda el cliente en ir desde el mostrador del primer empleado hasta el mostrador del segundo empleado está distribuido uniformemente entre 0.5 y 2.0 minutos.
- 5.10** Modifique el Problema 5.8 de modo que el modelo describa la siguiente característica del sistema: no existe cola frente al mostrador del segundo empleado. Si el primer empleado termina de atender a un cliente, y el segundo empleado se encuentra todavía ocupado, entonces el primer empleado debe permanecer con el cliente hasta que el segundo empleado quede libre. Esta característica se denomina “bloqueo”, ya que el cliente que permanece con el primer empleado, habiendo ya sido atendido, no recibe servicio, pero impide que un nuevo cliente pueda ser atendido por el primer empleado. Cuando el segundo empleado queda libre, el cliente abandona al primer empleado y pasa a ser atendido por el segundo. Entonces el primer empleado puede atender al primer cliente que se encuentra en su cola.
- 5.11** Suponga que en el sistema descrito en el Problema 5.8, existe una probabilidad igual a 0.2 de que un cliente que ha terminado de ser atendido por el segundo empleado se encuentre “insatisfecho”, y desee volver a ser atendido por ambos empleados. Estos clientes, una vez el segundo empleado ha finalizado de atenderles, en lugar de abandonar el sistema, se ponen de nuevo en la cola del primer empleado. Modifique el modelo de modo que describa esta característica del sistema.
- 5.12** Entre dos ciudades, A y B, existe un número fijo n de líneas telefónicas. Cada línea puede ser operada en cualquier dirección (es decir, puede soportar una llamada realizada de A a B, o de B a A), sin embargo, en un determinado instante, cada línea sólo soporta una única llamada.

Si una persona de la ciudad A o B desea realizar una llamada a la otra ciudad, y alguna de las n líneas se encuentra disponible, entonces ocupa una línea inmediatamente. Pero si todas las líneas se encuentran ocupadas, entonces la persona escucha una grabación que le indica que debe colgar e intentarlo más tarde.

El tiempo que transcurre entre intentos de llamada desde A a B está distribuido exponencialmente, con media 10 segundos, y desde B a A está distribuido exponencialmente con media 12 segundos. La duración de la conversación también está distribuida exponencialmente, con media 4 minutos, con independencia de la procedencia de la llamada.

Inicialmente todas las líneas se encuentran libres, y la simulación se ejecuta durante 12 horas. Se pretende calcular:

- El número medio de líneas que se encuentran ocupadas.
 - La proporción de llamadas que no pueden realizarse, por encontrarse todas las líneas ocupadas.
 - El número total de llamadas realizadas desde cada ciudad.
- 5.13** Una empresa tiene un sistema informático que consiste en una única CPU y n terminales. El operario que trabaja en cada terminal “piensa” durante un tiempo, que está distribuido exponencialmente con media 25 segundos, y entonces envía a la CPU un trabajo, que requiere un tiempo de servicio S , que está distribuido exponencialmente, con media 0.8 segundos.

Las tareas que llegan a la CPU forman una única cola, pero son atendidas de forma *round-robin*, en lugar que FIFO (ver la Figura 5.12). Esto es, la CPU asigna a cada tarea un tiempo de ejecución máximo de $q = 0.1$ minutos. Si q es mayor que el tiempo que resta para finalizar la ejecución del trabajo, s , entonces la CPU invierte $s + \tau$ segundos

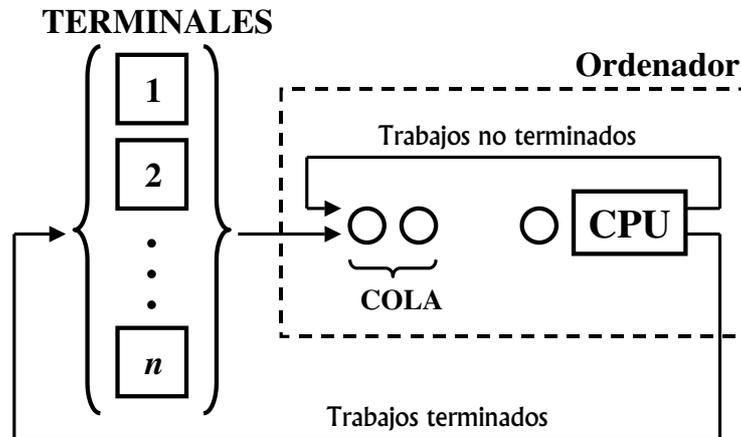


Figura 5.12: Modelo de un sistema informático de tiempo compartido.

en ejecutarlo y seguidamente lo devuelve al terminal. Para los trabajos recién llegados a la CPU se cumple $s := S$. τ es un tiempo fijo: $\tau = 0.015$ segundos.

Sin embargo, si el tiempo necesario para completar la ejecución de la tarea, s , es mayor que q , entonces la CPU invierte $q + \tau$ segundos en procesar el trabajo y a continuación pone el trabajo al final de la cola, decrementando el tiempo que resta para completarlo en q unidades ($s := s - q$). Este proceso se repite hasta que el trabajo es completado y devuelto al terminal.

El tiempo de respuesta del trabajo i , R_i , se define como el tiempo que transcurre desde que el terminal lo envía a la CPU y la CPU lo devuelve. El objetivo del estudio es determinar el número máximo de terminales que pueden conectarse en el sistema, de modo que el tiempo medio de respuesta no supere los 30 segundos. La condición de finalización de la simulación es que la CPU haya devuelto a los terminales 1000 trabajos. El tiempo medio de respuesta se calcula sobre estos 1000 trabajos.

Realizar el diagrama de flujo del modelo de simulación que posibilite llevar a cabo este estudio.

ANEXO 1

Programa de simulación de una cola atendida por un único empleado

Fichero cola.h (gestión de las colas)

```
#include <stdio.h>
#include <stdlib.h>
#include <alloc.h>

enum booleana {false,true};
typedef enum booleana booleana;

struct cola {
    float      dato;
    struct cola *siguiente;
};

struct gestion_cola {
    struct cola *inicio,
               *fin;
};

booleana Esta_vacia(g_cola)
struct gestion_cola *g_cola;
{
    if ( g_cola->inicio == NULL )
        return(true);
    else
        return(false);
}

booleana Anadir_elemento(g_cola,dato)
struct gestion_cola *g_cola;
float  dato;
{
    booleana error;
    if ( Esta_vacia(g_cola) == true )
        g_cola->inicio = g_cola->fin =
            (struct cola *) calloc(1,sizeof(struct cola));
    else
        g_cola->fin= (g_cola->fin)->siguiente =
            (struct cola *) calloc(1,sizeof(struct cola));
    if (g_cola->fin == NULL) {
        printf("\nERROR: No hay memoria disponible\n");
        error = true;
    }
    else {
        (g_cola->fin)->dato = dato;
        error = false;
    };
    return(error);
}

booleana Sacar_elemento(g_cola,dato)
struct gestion_cola *g_cola;
float  *dato;
{
    struct cola *aux;
    booleana error;
    if ( Esta_vacia(g_cola) == false ) {
        *dato = (g_cola->inicio)->dato;
        aux = g_cola->inicio;
        if ( (g_cola->inicio)->siguiente == NULL )
```

```
        gCola->inicio = gCola->fin = NULL;
    else
        gCola->inicio = (gCola->inicio)->siguiente;
    free(aux);
    error = false;
    }
else {
    printf("\nERROR: intento de sacar un elemento de una cola vacia\n");
    error = true;
    };
return(error);
}

void Imprime_cola(gCola)
struct gestion_cola *gCola;
{
    struct cola *aux;
    if ( Esta_vacia(gCola) == true ) {
        printf("Cola vacia");
    }
    else {
        aux = gCola->inicio;
        do {
            printf("%f\t",aux->dato);
            aux = aux->siguiente;
        } while ( aux != NULL );
    };
    printf("\n");
    return;
}
```

Programa principal y rutinas

```

#include <time.h>
#include <math.h>
#include <conio.h>
#include <dos.h>
#include "a:\cola.h"

#define NO_DEFINIDO -1.0

enum status_servidor {libre,ocupado};
typedef enum status_servidor status_servidor;

enum tipo_evento {llegada,marcha};
typedef enum tipo_evento tipo_evento;

struct estado_sistema {
    status_servidor      estado_empleado;
    unsigned int         n_clientes cola;
    struct gestion_cola  g_cola;
    float                t_evento_anterior;
};

struct lista_eventos {
    float    t_prox_llegada,
            t_marcha;
};

struct contad_estadisticos {
    unsigned int    n_clientes_inic_atencion;
    float          t_total_espera,
                integ_n_clientes_cola;
};

struct media_var_aleat {
    float    interv_llegadas,
            t_servicio;
};

float Distr_expon(media)
float    media;
{
    float    v_unif,
            v_expon;

    do {
        v_unif = rand() / 32767.0;
    } while (!(v_unif>0));
    v_expon = - media * log( v_unif );
    return(v_expon);
}

void Inicializacion (reloj,estado,eventos,cont_est,media)
float                *reloj;
struct media_var_aleat    media;
struct estado_sistema    *estado;
struct lista_eventos      *eventos;
struct contad_estadisticos *cont_est;
{
    /* Inicializacion del reloj de la simulacion */
    *reloj = 0;
    /* Inicializacion del estado del sistema */
    estado->estado_empleado = libre;
    estado->n_clientes_cola = 0;
    estado->t_evento_anterior = 0;
    (estado->g_cola).inicio = NULL;
    (estado->g_cola).fin = NULL;
    /* Inicializacion de la lista de eventos */
    eventos->t_prox_llegada = *reloj+Distr_expon(media.interv_llegadas);
}

```

```

eventos->t_marcha = NO_DEFINIDO;
/* Inicializacion de los contadores estadisticos */
cont_est->n_clientes_inic_atención = 0;
cont_est->t_total_espera = 0;
cont_est->integ_n_clientes cola = 0;
/* Inicializacion del generador de numeros aleatorios */
randomize();
return;
}

tipo_evento Rutina_tiempo (reloj,eventos)
float      *reloj;
struct lista_eventos  eventos;
{
    if ( eventos.t_prox_llegada < eventos.t_marcha ||
        eventos.t_marcha == NO_DEFINIDO ) {
        *reloj = eventos.t_prox_llegada;
        return(llegada);
    }
    else {
        *reloj = eventos.t_marcha;
        return(marcha);
    };
}

booleana Evento_llegada (reloj,estado,eventos,cont_est,media)
float      reloj;
struct estado_sistema      *estado;
struct lista_eventos      *eventos;
struct contad_estadisticos *cont_est;
struct media_var_aleat     media;
{
    booleana error = false;
    eventos->t_prox_llegada = reloj + Distr_expon( media.interv_llegadas );
    if (estado->estado_employado == libre) {
        estado->estado_employado = ocupado;
        cont_est->n_clientes_inic_atencion++;
        eventos->t_marcha = reloj + Distr_expon( media.t_servicio );
    }
    else {
        cont_est->integ_n_clientes_cola += (estado->n_clientes_cola)*
            (reloj-estado->t_evento_anterior);
        error = Anadir_elemento(&(estado->g_cola),reloj);
        (estado->n_clientes_cola)++;
    };
    estado->t_evento_anterior = reloj;
    return(error);
}

booleana Evento_marcha (reloj,estado,eventos,cont_est,media)
float      reloj;
struct media_var_aleat     media;
struct estado_sistema      *estado;
struct lista_eventos      *eventos;
struct contad_estadisticos *cont_est;
{
    float      *t_llego;
    booleana error = false;
    if ( estado->n_clientes_cola == 0 ) {
        estado->estado_employado = libre;
        eventos->t_marcha = NO_DEFINIDO;
    }
    else {
        eventos->t_marcha = reloj + Distr_expon(media.t_servicio);
        error = Sacar_elemento(&(estado->g_cola),t_llego);
        cont_est->t_total_espera += reloj-(*t_llego);
        cont_est->integ_n_clientes_cola += (estado->n_clientes_cola)*
            (reloj-estado->t_evento_anterior);
        (estado->n_clientes_cola)--;
    }
}

```

```

        cont_est->n_clientes_inic_atencion++;
    };
    estado->t_evento_anterior = reloj;
    return(error);
}

booleana Final_simulacion(n_clientes_inic_atencion,n_clientes_fin)
unsigned int    n_clientes_inic_atencion,
               n_clientes_fin;
{
    if ( n_clientes_inic_atencion < n_clientes_fin )
        return(false);
    else
        return(true);
}

void Entrada_datos (media_va,n_clientes_fin)
struct media_var_aleat    *media_va;
unsigned int              *n_clientes_fin;
{
    clrscr();
    printf("\tEntrada de datos:\n");
    printf("\t-----\n");
    printf("\nMedia del intervalo entre llegadas: ");
    scanf("%f",&(media_va->interv_llegadas));
    printf("\nMedia del tiempo de servicio: ");
    scanf("%f",&(media_va->t_servicio));
    printf("\nNum. clientes cuyo tiempo de espera debe simularse (1-65535): ");
    scanf("%d",n_clientes_fin);
    return;
}

void Informe (reloj,cont_est)
float          reloj;
struct contad_estadisticos    cont_est;
{
    float      t_espera_medio,
              nCola_medio;
    printf("\nFin de la simulacion...\n\n");
    printf("\tI N F O R M E\n");
    printf("\t-----\n\n");
    t_espera_medio = cont_est.t_total_espera / cont_est.n_clientes_inic_atencion;
    printf("\nTiempo medio de espera en la cola : %f\n\n",t_espera_medio);
    nCola_medio = cont_est.integ_n_clientesCola / reloj;
    printf("Numero medio de clientes en la cola: %f\n\n",nCola_medio);
    return;
}

void Imprime_estado_simulacion(reloj,estado,eventos,cont_est)
float          reloj;
struct estado_sistema    estado;
struct lista_eventos    eventos;
struct contad_estadisticos    cont_est;
{
    printf("-----\t\tReloj: %f\n",reloj);
    if (estado.estado_employado==libre)
        printf("Empleado LIBRE  ");
    else
        printf("Empleado OCUPADO");
    printf("\t\t%d clientes en la cola",estado.n_clientesCola);
    printf("\t\tUltimo evento: %f\nCola:\t",estado.t_evento_anterior);
    ImprimeCola(&(estado.gCola));
    printf("Lista de eventos: Llegada = %f  Marcha = ",
           eventos.t_prox_llegada);
    if (eventos.t_marcha==NO_DEFINIDO)
        printf("- \n");
    else
        printf("%f\n",eventos.t_marcha);
}

```

```
printf("Cont. estadisticos:\n");
printf("\tNumero de clientes que han empezado a ser atendidos:\t%d\n",
      cont_est.n_clientes_inic_atencion);
printf("\tTiempo total de espera de los clientes en la cola:\t%f\n",
      cont_est.t_total_espera);
printf("\tIntegral num. clientes en la cola respecto del tiempo:\t%f\n\n",
      cont_est.integ_n_clientes_cola);
}

main()
{
    struct media_var_aleat          media_va;
    struct estado_sistema          estado;
    struct lista_eventos           eventos;
    struct contad_estadisticos     cont_est;
    unsigned int                   n_clientes_fin;
    float                           reloj;
    booleana                       error;
    Entrada_datos(&media_va,&n_clientes_fin);
    Inicializacion(&reloj,&estado,&eventos,&cont_est,media_va);
    Imprime_estado_simulacion(reloj,estado,eventos,cont_est);
    do {
        if (Rutina_tiempo(&reloj,eventos)==llegada)
            error=Evento_llegada(reloj,&estado,&eventos,&cont_est,media_va);
        else
            error=Evento_marcha(reloj,&estado,&eventos,&cont_est,media_va);
        Imprime_estado_simulacion(reloj,estado,eventos,cont_est);
    } while (Final_simulacion(cont_est.n_clientes_inic_atencion,
                              n_clientes_fin) == false && error == false);
    if (error == false)
        Informe(reloj,cont_est);
    else
        printf("Terminacion anormal...\n");
    return;
}
```

ANEXO 2

Programa de simulación de la gestión del inventario de un producto

```
#include <stdio.h>
#include <stdlib.h>
#include <time.h>
#include <conio.h>
#include <dos.h>
#include <math.h>

#define NO_DEFINIDO -1.0

enum tipo_evento {
    llegada_prod_suministrador,
    cliente_realiza_pedido,
    fin_simulacion,
    evaluacion_inventario
};

typedef enum tipo_evento tipo_evento;

struct estado_sistema {
    int    inventario,
        prod_pedido_suministrador;
    float t_ultimo_cambio_inventario;
};

struct lista_eventos {
    float t_llegada_producto,
        t_cliente_pedido,
        t_fin_simulacion,
        t_evaluacion_inventario;
};

struct contad_estadisticos {
    float coste_total_pedido,
        integ_prod_inventario,
        integ_demanda_no_satisfecha;
};

struct v_a_discreta {
    float valor_variable,
        probabilidad_valor_variable;
};

struct politica_inventario {
    int    S,
        s;
};

struct precio_suministrador {
    float gestion,
        unidad_producto;
};

struct costes {
    struct precio_suministrador pedido;
    float    mantenimiento,
        deficit;
};

struct intervalo {
    float inicio,
        fin;
};
```

```

float Distr_expon(media)
float media;
{
    float v_unif,
          v_expon;
    do {
        v_unif=rand()/32727.0;
    } while (!(v_unif>0));
    v_expon=-media*log(v_unif);
    return(v_expon);
}

float Uniforme_interv_A_B(interv)
struct intervalo interv;
/* Generacion U(a,b) */
{
    float v_unif;
    v_unif=rand()/32727.0;
    return(interv.inicio+(interv.fin-interv.inicio)*v_unif);
}

float Var_aleat_discreta(var_aleat_discreta)
struct v_a_discreta *var_aleat_discreta;
{
    float v_unif,
          prob_acumulada=0;
    struct v_a_discreta *aux;
    v_unif=rand()/32727.0;
    if (v_unif>1.0) v_unif=1.0;
    aux=var_aleat_discreta;
    while (v_unif>(prob_acumulada+aux->probabilidad_valor_variable)) {
        prob_acumulada+=aux->probabilidad_valor_variable;
        aux++;
    }
    return(aux->valor_variable);
}

void Inicializacion (reloj,estado,eventos,cont_est,t_fin_simulacion,
media_interv_pedidos_cliente,interv_unif,
unidades_pedidas,coste)
float *reloj,
t_fin_simulacion,
*media_interv_pedidos_cliente;
struct estado_sistema *estado;
struct lista_eventos *eventos;
struct contad_estadisticos *cont_est;
struct v_a_discreta unidades_pedidas[];
struct intervalo *interv_unif;
struct costes *coste;
{
    /* Inicializacion del reloj de la simulacion */
    *reloj=0;
    /* Variable aleatoria "intervalo entre pedidos sucesivos de clientes" */
    *media_interv_pedidos_cliente=0.1;
    /* Variable aleatoria "cantidad de producto pedido por el cliente" */
    unidades_pedidas[0].valor_variable=1.0;
    unidades_pedidas[0].probabilidad_valor_variable=0.16667;
    unidades_pedidas[1].valor_variable=2.0;
    unidades_pedidas[1].probabilidad_valor_variable=0.33333;
    unidades_pedidas[2].valor_variable=3.0;
    unidades_pedidas[2].probabilidad_valor_variable=0.33333;
    unidades_pedidas[3].valor_variable=4.0;
    unidades_pedidas[3].probabilidad_valor_variable=0.16667;
    /* Variable aleatoria "tiempo de servicio del suministrador" */
    interv_unif->inicio=0.5;
    interv_unif->fin=1.0;
    /* Costes */
    coste->pedido.gestion=3200.0;
}

```

```

coste->pedido.unidad_producto=300.0;
coste->mantenimiento=100.0;
coste->deficit=500.0;
/* Inicializacion del estado del sistema */
estado->inventario=60;
estado->prod_pedido_suministrador=0;
estado->t_ultimo_cambio_inventario=0;
/* Inicializacion de la lista de eventos */
eventos->t_llegada_producto=NO_DEFINIDO;
eventos->t_cliente_pedido=*reloj+Distr_expon(*media_interv_pedidos_cliente);
eventos->t_fin_simulacion=t_fin_simulacion;
eventos->t_evaluacion_inventario=0;
/* Inicializacion de los contadores estadisticos */
cont_est->coste_total_pedido=0;
cont_est->integ_prod_inventario=0;
cont_est->integ_demanda_no_satisfecha=0;
/* Inicializacion del generador de numeros aleatorios */
randomize();
return;
}

tipo_evento Rutina_tiempo (reloj,eventos)
float *reloj;
struct lista_eventos eventos;
{
    if ( ( eventos.t_fin_simulacion <= eventos.t_llegada_producto ||
          eventos.t_llegada_producto == NO_DEFINIDO )
        &&
        eventos.t_fin_simulacion <= eventos.t_cliente_pedido
        &&
        eventos.t_fin_simulacion <= eventos.t_evaluacion_inventario) {
        *reloj=eventos.t_fin_simulacion;
        return(fin_simulacion);
    }
    else {
        if ( eventos.t_llegada_producto != NO_DEFINIDO
            &&
            eventos.t_llegada_producto < eventos.t_fin_simulacion
            &&
            eventos.t_llegada_producto <= eventos.t_cliente_pedido
            &&
            eventos.t_llegada_producto <= eventos.t_evaluacion_inventario) {
            *reloj=eventos.t_llegada_producto;
            return(llegada_prod_suministrador);
        }
        else {
            if (eventos.t_cliente_pedido < eventos.t_fin_simulacion
                &&
                (eventos.t_cliente_pedido < eventos.t_llegada_producto ||
                 eventos.t_llegada_producto == NO_DEFINIDO)
                &&
                eventos.t_cliente_pedido<=eventos.t_evaluacion_inventario
            ) {
                *reloj=eventos.t_cliente_pedido;
                return(cliente_realiza_pedido);
            }
            else {
                *reloj=eventos.t_evaluacion_inventario;
                return(evaluacion_inventario);
            }
        }
    }
};

void Actualiza_integrales(reloj,estado,cont_est)
float reloj;
struct estado_sistema *estado;
struct contad_estadisticos *cont_est;
{
    if (estado->inventario>0)
        cont_est->integ_prod_inventario+=estado->inventario*
            (reloj-estado->t_ultimo_cambio_inventario);
    if (estado->inventario<0)
        cont_est->integ_demanda_no_satisfecha+=-estado->inventario*
            (reloj-estado->t_ultimo_cambio_inventario);
    estado->t_ultimo_cambio_inventario=reloj;
    return;
}

```

```

}

void Evento_fin_simulacion (reloj,estado,cont_est)
float reloj;
struct estado_sistema *estado;
struct contad_estadisticos *cont_est;
{
    Actualiza_integrales(reloj,estado,cont_est);
    return;
}

void Evento_llegada_producto (reloj,estado,eventos,cont_est)
float reloj;
struct estado_sistema *estado;
struct lista_eventos *eventos;
struct contad_estadisticos *cont_est;
{
    Actualiza_integrales(reloj,estado,cont_est);
    estado->inventario+=estado->prod_pedido_suministrador;
    estado->prod_pedido_suministrador=0;
    eventos->t_llegada_producto=NO_DEFINIDO;
    return;
}

void Evento_pedido_cliente (reloj,estado,eventos,cont_est,
distr_num_prod_demandados,media_intervalo_pedidos_cliente)
float
    reloj,
    media_intervalo_pedidos_cliente;
struct estado_sistema *estado;
struct lista_eventos *eventos;
struct contad_estadisticos *cont_est;
struct v_a_discreta *distr_num_prod_demandados;
{
    Actualiza_integrales(reloj,estado,cont_est);
    estado->inventario -= (int) Var_aleat_discreta(distr_num_prod_demandados);
    eventos->t_cliente_pedido=reloj+
        Distr_expon(media_intervalo_pedidos_cliente);
    return;
}

void Evento_evaluacion_inventario(reloj,estado,eventos,cont_est,
polit_invent,coste,interv_v_a)
float
    reloj;
struct estado_sistema *estado;
struct lista_eventos *eventos;
struct contad_estadisticos *cont_est;
struct politica_inventario polit_invent;
struct costes coste;
struct intervalo interv_v_a;
{
    int pedido;
    if (estado->inventario < polit_invent.s) {
        pedido=polit_invent.S-estado->inventario;
        if (pedido>0) {
            estado->prod_pedido_suministrador=pedido;
            cont_est->coste_total_pedido+=coste.pedido.gestion+
                pedido*coste.pedido.unidad_producto;
            eventos->t_llegada_producto=reloj+
                Uniforme_interv_A_B(interv_v_a);
        };
    };
    (eventos->t_evaluacion_inventario)++;
    return;
}

void Entrada_datos (polit_invent,t_fin_simulacion)
struct politica_inventario *polit_invent;

```

```

float                                *t_fin_simulacion;
{
    clrscr();
    printf("\tEntrada de datos:\n");
    printf("\t-----\n");
    printf("\nParametros de gestion del inventario:\n\n\tParametro s: ");
    scanf("%d",&(polit_invent->s));
    printf("\n\tParametro S: ");
    scanf("%d",&(polit_invent->S));
    printf("\nDuracion de la simulacion (meses): ");
    scanf("%f",t_fin_simulacion);
    return;
}

void Informe (reloj,cont_est,coste)
float                                reloj;
struct contad_estadisticos           cont_est;
struct costes                         coste;
{
    float                                media_pedido,
                                          media_mantenimiento,
                                          media_deficit,
                                          media_coste_total;
    printf("\nFin de la simulacion...\n\n");
    printf("\tI N F O R M E\n");
    printf("\t-----\n\n");
    printf("\tCoste medio mensual de:\n\n");
    media_pedido=cont_est.coste_total_pedido/reloj;
    printf("\tPedidos: \t\t\t%f\n",media_pedido);
    media_mantenimiento=cont_est.integ_prod_inventario*coste.mantenimiento/reloj;
    printf("\tMantenimiento de inventario: \t%f\n",media_mantenimiento);
    media_deficit=cont_est.integ_demanda_no_satisfecha*coste.deficit/reloj;
    printf("\tDeficit de inventario: \t\t%f\n",media_deficit);
    media_coste_total=media_pedido+media_mantenimiento+media_deficit;
    printf("\tCoste total: \t\t\t%f\n",media_coste_total);
    return;
}

void Imprime_estado_simulacion(reloj,estado,eventos,cont_est)
float                                reloj;
struct estado_sistema                 estado;
struct lista_eventos                  eventos;
struct contad_estadisticos           cont_est;
{
    printf("-----\t\tReloj: %f\n",reloj);
    printf("Inventario: %d\t\tProducto encargado al suministrador: %d\n",
           estado.inventario,estado.prod_pedido_suministrador);
    printf("Instante ultimo cambio en el inventario: %f\n",
           estado.t_ultimo_cambio_inventario);
    printf("Lista de eventos: \n");
    printf("Pedido de un cliente: %f\tLlegada de producto: ",
           eventos.t_cliente_pedido);
    if (eventos.t_llegada_producto==NO_DEFINIDO)
        printf("- \n");
    else
        printf("%f\n",eventos.t_llegada_producto);
    printf("Fin de la simulacion: %f\tEvaluacion inventario: %f\n",
           eventos.t_fin_simulacion,eventos.t_evaluacion_inventario);
    printf("Cont. estadisticos:\n");
    printf("Coste pedido: %f\tIntegral I+: %f\tIntegral I-: %f\n\n",
           cont_est.coste_total_pedido,cont_est.integ_prod_inventario,
           cont_est.integ_demanda_no_satisfecha);
}

main()
{
    struct estado_sistema                 estado;
    struct lista_eventos                  eventos;
    struct contad_estadisticos           cont_est;

```

```
struct politica_inventario      polit_invent;
struct intervalo                interv_unif;
struct v_a_discreta            unidades_pedidas[4];
struct costes                   coste;
float                           reloj,
                                t_fin_simulacion,
                                media_interv_pedidos_cliente;

tipo_evento                    event;
Entrada_datos(&polit_invent, &t_fin_simulacion);
Inicializacion(&reloj, &estado, &eventos, &cont_est, t_fin_simulacion,
              &media_interv_pedidos_cliente, &interv_unif,
              unidades_pedidas, &coste);
Imprime_estado_simulacion(reloj, estado, eventos, cont_est);
do {
    event=Rutina_tiempo(&reloj, eventos);
    if (event==llegada_prod_suministrador)
        Evento_llegada_producto(reloj, &estado, &eventos, &cont_est);
    else if (event==cliente_realiza_pedido)
        Evento_pedido_cliente(reloj, &estado, &eventos, &cont_est,
                               unidades_pedidas, media_interv_pedidos_cliente);
    else if (event==fin_simulacion)
        Evento_fin_simulacion(reloj, &estado, &cont_est);
    else
        Evento_evaluacion_inventario(reloj, &estado, &eventos, &cont_est,
                                      polit_invent, coste, interv_unif);
    Imprime_estado_simulacion(reloj, estado, eventos, cont_est);
}
while (event!=fin_simulacion);
Informe(reloj, cont_est, coste);
return;
}
```

Tema 6

Simulación usando Arena

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los fundamentos del modelado orientado a los procesos.
- Diseñar y realizar evaluaciones básicas de modelos dinámicos para simulación por ordenador.
- Diseñar, codificar usando Arena, comprobar y depurar programas de simulación de eventos discretos, de tiempo continuo e híbridos.

6.1. Introducción

En este tema se presentan los conceptos básicos del *modelado orientado a los procesos*. De acuerdo con esta metodología, el modelo de un sistema se realiza estudiando las *entidades* que circulan a través del sistema.

Para ilustrar los conceptos expuestos, se va a emplear el entorno de modelado Arena. Como guía en el manejo de Arena se recomienda la consulta de dos manuales en formato pdf almacenados en el directorio creado al instalar Arena. Los manuales son:

- Arena Standard Edition, User's Guide.
- Variables Guide.

A lo largo del tema se hará referencia a ellos, guiando la lectura del alumno. Aparte de esta documentación, Arena dispone de una potente herramienta de ayuda. Conviene que el alumno se familiarice cuanto antes con su manejo.

El capítulo "Getting Started" de la guía "Arena Standard Edition, User's Guide" es una introducción muy didáctica al modelado, la simulación y la visualización de los resultados usando Arena. Las explicaciones se apoyan en un modelo sencillo, que es descrito en el siguiente ejemplo.

Ejemplo 6.1.1. *Se pretende estudiar cierto proceso de recepción y revisión de solicitudes de créditos hipotecarios ("home mortgage application process"). La revisión de las solicitudes es realizada por un único empleado, y consiste en decidir si la solicitud de crédito contiene toda la documentación requerida o si, por el contrario, falta algún documento y debe ser rechazada.*

El objetivo del estudio de simulación es responder a las siguientes cinco preguntas:

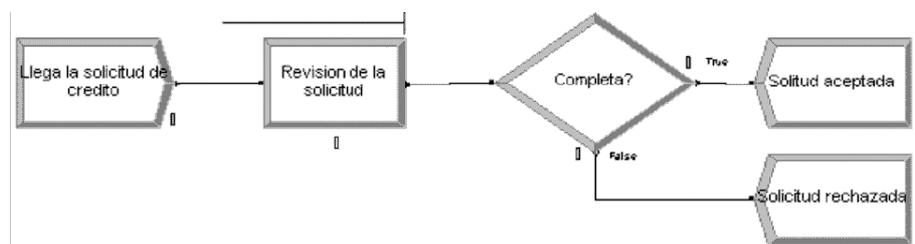


Figura 6.1: Revisión de las solicitudes de créditos hipotecario.

- ¿Cuánto tiempo transcurre, en promedio, desde que una solicitud es recibida hasta que finaliza su revisión?
- Dado que el sueldo del empleado es 12 dólares/hora, ¿Cuál es el coste promedio de revisar cada solicitud?
- ¿Cuál es el tiempo máximo de revisión de una solicitud?
- ¿Cuál es el número máximo de solicitudes en espera de ser revisadas?
- ¿Qué proporción del tiempo el empleado se encuentra ocupado?

El modelo contiene un único tipo de entidad: la solicitud de la hipoteca. Existe un único proceso, el de revisión de las solicitudes, que se lleva a cabo usando un único recurso: el empleado. En la Figura 6.1 se muestra el diagrama de flujo del modelo.

El modelo contiene tres elementos de incertidumbre, que son modelados de la forma siguiente:

- Intervalo de tiempo entre llegadas consecutivas de solicitudes:

$$\sim \text{expo}(2 \text{ horas})$$

- Tiempo que tarda el empleado en revisar una solicitud:

$$\sim \text{triang}(1 \text{ hora}, 1.75 \text{ horas}, 3 \text{ horas})$$

- La decisión de aceptar o devolver una solicitud de basa en si está completa o incompleta. Se define una probabilidad de que la solicitud esté completa (0.88) y la complementaria (0.12) de que esté incompleta. Por consiguiente, este atributo de la entidad "Solicitud" se modela mediante una variable aleatoria discreta.

Las características de las distribuciones de probabilidad exponencial y triangular, así como de las demás distribuciones que soporta Arena, están resumidas en el apéndice "Statistical Distributions" de la guía "Arena Standard Edition, User's Guide".

Al comenzar la jornada de trabajo el empleado tiene pendientes todas las solicitudes que no revisó los días anteriores. Por consiguiente, la finalización de la jornada laboral no supone un evento relevante: el empleado reanuda el trabajo en el punto en que lo dejó el día anterior. Por este motivo, y puesto que el coste se calcula por hora, al diseñar el experimento puede suponerse que el sistema funciona ininterrumpidamente. En concreto, se simula el funcionamiento del sistema durante 20 días, suponiendo que funciona 24 horas al día. Dado que se trata de un ejemplo introductorio, se realiza una única réplica de la simulación. □

Se sugiere al alumno que lea ahora el capítulo “Getting Started” de la guía “Arena Standard Edition, User’s Guide” y que realice por sí mismo el ejemplo siguiendo los pasos descritos en el manual.

Durante la construcción del modelo, es muy conveniente ir consultando el capítulo “The Basic Process Panel” de la guía, en particular los apartados referentes a los módulos de flujo (“flowchart modules”) “Create”, “Process”, “Decide” y “Dispose”, y a los módulos de datos (“data modules”) “Entity”, “Queue” y “Resource”.

6.2. Modelado de operaciones básicas

En esta sección se introducen algunos de los conceptos básicos del modelado y la simulación usando Arena. En concreto, se discute el tipo de modelado que puede hacerse empleando los módulos del panel “Basic Process”.

Módulos de flujo y de datos

La descripción del modelo en Arena se hace por medio de módulos, que se encuentran organizados en diferentes *paneles*. Pueden diferenciarse las dos siguientes familias de módulos (Kelton, Sadowski & Sadowski 2002):

- *Módulos de flujo*. Permiten describir el flujo de las entidades a través del sistema. El *diagrama de módulos* que representa el flujo de las entidades se construye pinchando los correspondientes módulos de flujo y arrastrándolos desde su panel hasta la ventana del modelo. Haciendo doble “clic” sobre uno de estos módulos en la ventana del modelo se accede a sus parámetros.

Los módulos de flujo del panel “Basic Process” permiten describir:

- Los puntos de entrada de las entidades al sistema (módulo “Create”).
 - Los procesos (módulo “Process”).
 - Las bifurcaciones en el flujo de las entidades (módulo “Decide”).
 - La formación de grupos de entidades de un tamaño establecido (módulo “Batch”).
 - La duplicación de las entidades para su procesamiento en paralelo, o la separación de las entidades previamente agrupadas (módulo “Separate”).
 - Cambios en el valor de algún parámetro durante la simulación, tales como el tipo de una entidad o una variable del modelo (módulo “Assign”).
 - El cálculo de estadísticos, tales como el número total de entidades y el tiempo de ciclo (módulo “Record”).
 - Los puntos por los que las entidades abandonan el sistema (módulo “Dispose”).
- *Módulos de datos*. Presentan la información acerca de los diferentes aspectos del modelo de forma estructurada, de modo que su consulta y modificación se facilita enormemente. Por ejemplo, los módulos de datos “Entity” (Entidad), “Queue” (Cola) y “Resource” (Recurso) contienen las características de todos los tipos de entidad, colas y recursos definidos en el modelo. Asimismo, existen módulos de datos que complementan la descripción del modelo proporcionada por los módulos de flujo. Por ejemplo, el módulo de datos “Schedule” (Planificación) permite describir la variación temporal de la capacidad de un recurso.

Los módulos de datos no deben ser arrastrados a la ventana del modelo: en cada modelo sólo existe un módulo de datos de cada clase y está ubicado en su correspondiente panel. Para editar un módulo de datos basta con seleccionarlo, es decir, hacer “clic” sobre él: aparecerá la información en un formato similar al de una hoja de cálculo, pudiendo ser modificada, ampliada, etc.

Definición de los procesos

En el panel “Basic Process” se encuentra el módulo “Process”, que permite modelar los procesos. Para ello, es preciso pinchar el módulo en el panel y arrastrarlo a la ventana del modelo. A continuación, haciendo doble “clic” sobre el módulo así creado en la ventana del modelo, se accede a sus propiedades, que son las siguientes:

- La *acción* que realiza la entidad en el proceso. Existen cuatro posibles tipos de acción (“Delay”, “Seize-Delay”, “Seize-Delay-Release” y “Delay-Release”) y debe seleccionarse una de ellas.
 - “Delay”. La entidad permanece durante cierto tiempo esperando (Delay), no realizándose ninguna acción concreta sobre ella. Este tipo de proceso no requiere de recurso, y por ello el proceso puede ser capturado por varias entidades simultáneamente. Asimismo, dado que la finalidad de la espera no es aguardar a que el recurso sea liberado, también puede suceder que varias entidades finalicen la espera simultáneamente.
 - “Seize-Delay”. Si el recurso está ocupado, la entidad espera (Wait) en la cola hasta que el recurso quede disponible. Cuando el recurso está disponible, la entidad lo captura (Seize) y entonces espera (Delay) mientras el recurso realiza las operaciones pertinentes sobre ella, concluidas las cuales la entidad abandona el proceso, pero sin liberar el recurso.
 - “Seize-Delay-Release”. Cuando el recurso queda disponible la entidad lo captura (Seize), espera (Delay) mientras el recurso realiza las operaciones sobre ella, y a continuación libera (Release) el recurso y abandona el proceso.
 - “Delay-Release”. En este caso se supone que cuando la entidad entra en el proceso ya tiene capturado el recurso. Justo al entrar en el proceso, la entidad espera (Delay) mientras el recurso opera sobre ella, y a continuación libera (Release) el recurso, saliendo del proceso.
- Cuando la acción requiera de un recurso, deben definirse las *características del recurso*:
 - El nombre del recurso.
 - La cantidad (“Quantity”) de la capacidad del recurso que captura o libera la entidad en la acción.

Al definirse un nuevo recurso en un módulo “Process”, aquel se añade automáticamente a la lista de recursos del modelo, que se guarda en el módulo de datos “Resource”. Es en este módulo de datos donde deben definirse las propiedades del recurso, tales como su capacidad.

- La distribución de probabilidad del tiempo que dura la fase “Delay” de la acción. Asimismo, debe indicarse bajo qué concepto (“Allocation”) se desea que Arena contabilice el tiempo invertido en dicha fase. Existen las siguientes cinco posibilidades:
 - Value Added.
 - Non-Value Added.
 - Transfer.
 - Wait.
 - Other.

A la hora de computar el tiempo transcurrido por las entidades en la fase “Delay” de la acción, Arena contabiliza por separado los tiempos dedicados a cada uno de los cinco conceptos anteriores. Por ejemplo, el tiempo de ciclo de cada tipo de entidad se desglosa en la suma de las cinco variables siguientes: *VATime*, *NVATime*, *WaitTime*, *TranTime* y *OtherTime*. Estas variables almacenan el tiempo total acumulado en la fase “Delay” en procesos designados como “Value Added”, “Non-Value Added”, “Wait”, “Transfer” y “Other” respectivamente. Asimismo, se define la variable tiempo de ciclo, “TotalTime”, que es la suma de los cinco tiempos anteriores. Al concepto “Wait” se asigna también el tiempo que pasan las entidades en las colas de los recursos. Como se discutirá más adelante, la asignación del tiempo a los diferentes conceptos tiene especial relevancia cuando se desean realizar estimaciones del coste, ya que Arena computa separadamente el coste del tiempo invertido en cada concepto.

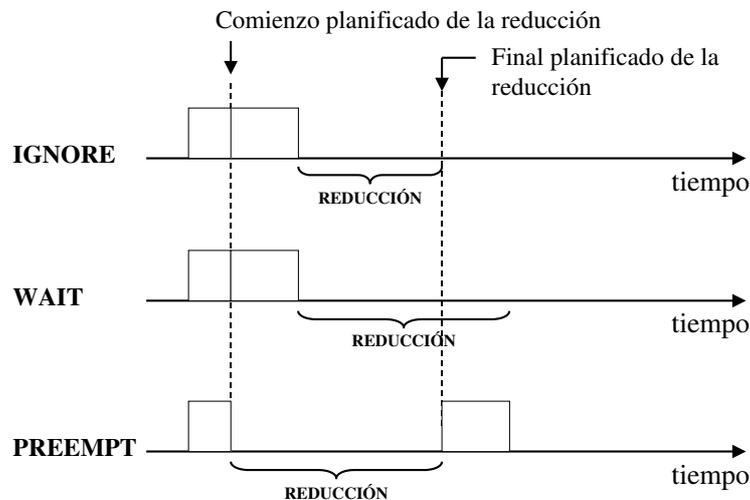


Figura 6.2: Opciones "Ignore", "Wait" y "Preempt".

Planificación de la capacidad de los recursos

En el módulo de datos "Resource", Arena lleva automáticamente registro de todos los recursos que se han definido en el modelo. En este módulo de datos deben definirse las propiedades de los recursos, entre las que se encuentra la capacidad (Capacity) del recurso.

El contenido de la columna "Type" del módulo "Resource" determina si el recurso mantiene una capacidad constante durante toda la simulación ("*Fixed Capacity*"), o si, por el contrario, la capacidad varía en el tiempo de acuerdo a determinada planificación ("*Based on Schedule*"). En este último caso, debe proporcionarse la siguiente información:

- El nombre de la planificación ("*Schedule Name*").
- La regla ("*Schedule Rule*") que determina cómo debe producirse la reducción en la capacidad. Existen las tres posibles reglas siguientes: "*Wait*", "*Ignore*" y "*Preempt*".
- La evolución de la capacidad del recurso en función del tiempo. Se define en el módulo de datos "Schedule".

Al asignar nombre a una nueva planificación en el módulo de datos "Resource", automáticamente Arena añade esa nueva planificación en el módulo de datos "Schedule", que es donde se lleva registro de las planificaciones definidas en el modelo y de las propiedades de aquellas. En concreto, es en el módulo "Schedule" donde se define la planificación en sí, es decir, donde se especifica cuál es la evolución temporal de la capacidad. El hecho de definir la planificación separadamente de la definición del recurso facilita la definición del modelo en el caso en el cual haya varios recursos que tengan la misma planificación de la capacidad: la planificación se define una única vez, en el módulo "Schedule", y se referencia por su nombre tantas veces como sea preciso desde el módulo "Resource".

Las reglas ("*Schedule Rules*") funcionan de la forma siguiente. Si está planificada una reducción en la capacidad y ésta se encuentra libre, cualquiera de estas tres reglas hace que pase al estado "Inactive" (Inactivo). Por el contrario, si en el instante en que está planificada la reducción en la capacidad, una entidad tiene capturado el recurso, la acción depende de la regla elegida (ver la Figura 6.2):

- La regla "*Ignore*" espera a que la entidad libere el recurso para poner éste en el estado "Inactive". Sin embargo, el tiempo que debe durar la reducción en la capacidad comienza a contabilizarse a partir del instante en que teóricamente debería haberse producido, no a partir del instante en que efectivamente se produce. Es decir, el tiempo real durante el

cual se reduce la capacidad es igual al tiempo de reducción planificado, menos el tiempo que transcurre hasta que la entidad libera el recurso. Si la entidad no libera el recurso durante todo el tiempo en que estaba planificada la reducción, entonces la reducción no se produce. Aplicando esta regla, pueden no coincidir con los valores planificados tanto el tiempo de inicio de la reducción como su duración.

- La regla *“Wait”* espera a reducir la capacidad del recurso a que la entidad lo libere. Una vez el recurso queda libre, éste pasa al estado *“Inactive”* y así permanece durante el total del tiempo establecido para la reducción en la capacidad. Por tanto, el tiempo durante el cual se reduce la capacidad siempre coincide con el planificado, pero no así el instante de inicio de la reducción, que puede retrasarse.
- La regla *“Preempt”* interrumpe la acción que está realizando el recurso sobre la entidad y hace que el recurso pase al estado *“Inactive”*. Entonces, la entidad queda a la espera de que transcurra el tiempo de reducción de la capacidad y, una vez transcurrido, vuelve a capturar el recurso, reanudándose el proceso en el punto en que se interrumpió. Esta regla satisface de forma precisa tanto el instante de inicio de la reducción como su duración.

Debe examinarse el sistema y escoger la regla que mejor describe lo que sucede en el sistema en el instante de cambio en la capacidad o fallo de un recurso. Si el recurso en cuestión es el cuello de botella del sistema, la elección de la regla puede afectar significativamente a los resultados obtenidos.

Existen algunas orientaciones prácticas acerca de en qué situaciones usar estas reglas (Kelton et al. 2002):

- Si el tiempo de la reducción de la capacidad es muy largo comparado con el tiempo de proceso, lo más adecuado puede ser usar la regla *“Ignore”*.
- Si el intervalo entre las reducciones en la capacidad es largo comparado con el tiempo de reducción en sí, es adecuado usar la regla *“Wait”*.

Fallos en los recursos

Arena permite modelar fallos en los recursos. De hecho, un recurso puede encontrarse en cada instante de la simulación en uno de los siguientes estados:

- *Libre* (*“Idle”*). El recurso se encuentra listo para ser capturado por una entidad, sin embargo ninguna de ellas lo ha hecho.
- *Ocupado* (*“Busy”*). El recurso se encuentra actualmente capturado por una entidad.
- *Inactivo* (*“Inactive”*). Mientras la capacidad del recurso es cero.
- *Fuera de servicio* (*“Failed”*). Cuando debido a una avería, un mantenimiento, etc. el recurso no se encuentra disponible para ser capturado por ninguna entidad.

Arena contabiliza el tiempo que cada recurso pasa en cada uno de estos cuatro estados, e incluye por defecto este dato en el informe del resultado de la simulación.

Para modelar los fallos en un recurso, en primer lugar es preciso definir el tipo de fallo, y, a continuación, se asocia el tipo de fallo con el recurso. De este modo, es posible asociar a varios recursos el mismo tipo de fallo. En estos casos, Arena genera secuencias de números aleatorios independientes para simular los fallos de cada recurso.

La clase de fallo puede definirse en el módulo de datos *“Failure”*, que está en el panel *“Advanced Process”*, y consiste en asignarle un nombre, que puede escogerse libremente, y definir el tipo: *“Count”* o *“Time”*.

- *Count* (cuenta) hace que el recurso falle después de procesar un determinado número de entidades, bien fijo o bien calculado de una expresión. Este tipo de fallo permite modelar mantenimientos preventivos (limpieza, sustitución de piezas, ajustes, etc.) realizados a la máquina cada cierto número de entidades procesadas. Aunque una parada por mantenimiento no es un fallo de la máquina propiamente dicho, en este caso se considera como tal, ya que evita que el recurso pueda ser capturado por las entidades.

Debe definirse la forma de calcular el número de entidades (“Count”) necesario para que se produzca el fallo. Este número puede ser fijo, estar distribuido aleatoriamente o ser calculado de una expresión. Asimismo, debe definirse la forma de calcular el tiempo durante el cual el recurso permanece fuera de servicio (“Down Time”).

- *Time* hace que la máquina falle de acuerdo a un criterio basado en el tiempo. En este caso, debe indicarse cómo calcular el tiempo entre fallos (“Up Time”) y el tiempo durante el cual el recurso está fuera de servicio (“Down Time”). También debe definirse (en la columna “Uptime in this State only”) en qué estado del recurso se considera que la máquina está “Up”. Por ejemplo, si el mecanismo de fallo es el desgaste de la maquinaria del recurso o el agotamiento de un consumible, sólo debe contabilizarse como “Up Time” el tiempo durante el cual el recurso está ocupado. Por defecto, se contabiliza el tiempo que pasa la máquina tanto en el estado libre como ocupado.

Una vez definido el tipo de fallo, éste debe asociarse al recurso. Para ello, debe abrirse el módulo de datos “Resource” (haciendo “clic” sobre él, en el panel “Basic Process”), que contiene la información de todos los recursos del modelo. En la columna “Failures”, de la fila correspondiente al recurso, debe seleccionarse el nombre del tipo de fallo. Asimismo, debe definirse la regla a seguir al poner el recurso fuera de servicio: “Wait”, “Ignore” o “Preempt”. El significado de estas reglas es el mismo que en el caso de las planificaciones.

Definición del experimento

Pulsando *Run / Setup* ... se abre una ventana de diálogo en la cual se introduce la información relativa al experimento. La ventana posee varias lengüetas (“Project Parameters”, “Replication Parameters”, etc.), que permiten configurar los diferentes aspectos del experimento.

Seleccionando la lengüeta “Project Parameters” puede definirse:

- El nombre del proyecto y del analista.
- Sobre qué aspectos del modelo (costes, entidades, recursos, colas, etc.) se desea que Arena vuelque información en el fichero de salida de los resultados de la simulación.

Seleccionando la lengüeta “Replication Parameters” puede indicarse:

- El número de réplicas independientes (“Number of Replications”) que se desean realizar de la simulación.
- Cuando se realiza más de una réplica, es preciso indicar cómo debe procederse entre las réplicas en lo que respecta al estado del sistema y a los informes. Para ello deben seleccionarse o no las casillas correspondientes a “Statistics” y “System” en la parte de la ventana “Initialize Between Replications”. Existen cuatro posibilidades:

- *Initialize System (sí), Initialize Statistics (sí)*.

Se realizan réplicas estadísticamente independientes de la simulación (es decir, se emplea una secuencia de números aleatorios diferente para cada réplica), inicializándose el sistema entre réplicas. Se obtienen informes independientes de los resultados obtenidos en cada réplica.

- *Initialize System (sí), Initialize Statistics (no)*.

Se realizan réplicas estadísticamente independientes, inicializándose el sistema entre ellas. Los informes son acumulativos. Es decir, el primer informe contiene los resultados de la primera réplica. El segundo informe los resultados acumulados de la primera y la segunda réplica, y así sucesivamente.

- *Initialize System (no), Initialize Statistics (sí)*.

Se realizan las réplicas de modo que el estado final del sistema en cada réplica se toma como el estado inicial en la siguiente. El sistema no se inicializa entre réplicas (por ejemplo, el tiempo continua avanzando réplica tras réplica), con lo cual las réplicas no son necesariamente independientes ni idénticamente distribuidas en el sentido estadístico estricto, lo cual tiene un impacto en cómo debe realizarse el análisis estadístico de los resultados.

Los informes no son acumulativos, sino que Arena genera un informe para cada réplica.

- *Initialize System (no), Initialize Statistics (no)*.

Se realizan réplicas de la simulación sin inicializar el estado del sistema entre ellas. Los informes son acumulativos. Por ejemplo, el décimo informe sería el mismo que se hubiera obtenido si se hubiera realizado una única réplica de longitud igual a la longitud acumulada de las 10 primeras réplicas.

- Si se desea que el sistema opere durante cierto tiempo (“Warm-up Period”) previamente al inicio del experimento, de modo que el sistema al iniciar el experimento no se encuentre vacío de entidades, debe especificarse la duración de este periodo.
- La unidad de tiempo (segundos, minutos, horas o días) que debe emplear internamente Arena al realizar la simulación debe especificarse en la casilla “Base Time Units”. Es conveniente escoger la unidad de tiempo que más se asemeje a las constantes de tiempo del sistema.
- La duración de la simulación (“Replication Length”) o su condición de finalización (“Terminating Condition”).

Informe de los resultados de la simulación

Los resultados de la ejecución de la simulación son accesibles desde el panel “Reports”, de la Barra del Proyecto. Los datos mostrados en estos informes son extraídos de un fichero Access (con extensión *.mdb*) creado durante la simulación con este fin. Estos mismos resultados están igualmente recogidos, aunque de forma más compacta, en el informe resumen de SIMAN (fichero *ascii* con extensión *.out*). Estos dos ficheros de salida de la simulación, *.mdb* y *.out*, son generados en el mismo directorio donde se guarda el modelo que ha sido simulado (fichero con extensión *.doe*).

En el informe del fichero *.out*, Arena muestra los estadísticos agrupados de la forma siguiente:

- “*Tally statistics*” (Estadísticos de tiempo discreto). Son aquellos que se obtienen de realizar un cálculo (por ejemplo, la media, el máximo o el mínimo) sobre una secuencia discreta de observaciones.
- “*Counter statistics*” (Contadores estadísticos). Son sumas que acumulan los sucesivos valores de algo. Por ejemplo, el número de solicitudes recibidas o que comienzan a ser revisadas, el tiempo total de espera de las solicitudes, el tiempo total que el empleado pasa ocupado, etc.
- “*Time-persistent statistics*” (Estadísticos de tiempo continuo). Son calculados al operar funciones de tiempo continuo, tales como acumuladores estadísticos que calculan la integral de una curva. Por ejemplo, la proporción de tiempo que un recurso está ocupado, el promedio de entidades en una cola, etc.
- “*Outputs*” (Salidas). Este tipo de estadístico está disponible cuando se ejecutan múltiples réplicas de la simulación.

Arena calcula por defecto determinados estadísticos¹, tales como:

- *Tally*: tiempos en cada proceso, tiempos en cada cola, etc.
- *Time-persistent*: número de entidades en cada cola, uso de cada recurso (“resource usage”) y utilización de cada recurso (“resource utilization”).
- *Counters*: tiempo acumulado, número total de entidades que llegan al sistema, número total de entidades que lo abandonan, etc.

El usuario puede habilitar o deshabilitar el cálculo de estos estadísticos al definir el experimento, seleccionando aquellos aspectos del modelo acerca de los cuales desea obtener información (*Run / Setup ... / Project Parameters*): coste, colas, recursos, entidades, etc.

¹El significado de las variables predefinidas de Arena puede consultarse en la guía “Arena Variables Guide”, que está disponible en formato *.pdf* en el directorio donde se ha instalado Arena.

Asimismo, el usuario puede definir el cálculo de otro tipo de magnitudes, empleando los módulos “Record”, situado en el panel “Basic Process”.

Para cada estadístico de tiempo discreto o continuo, Arena calcula el promedio por entidad (“Average”), el valor máximo (“Maximum Value”) y el mínimo (“Minimum Value”). Además, al final de cada réplica, Arena trata de calcular el intervalo del 95% de confianza del valor esperado en el estacionario del estadístico, usando el método de las *medias por lotes*. La columna “Half Width” contiene la mitad de este intervalo (es decir, el intervalo es la media mas - menos esa cantidad).

Para realizar la estimación del intervalo, Arena en primer lugar comprueba que el número de datos recogidos es suficiente para poder aplicar el método. Si no lo es, se muestra el mensaje “Insufficient”. Si hay suficientes datos, se hace el test sobre la correlación de los lotes de datos: para poder aplicar el método, los lotes no deben estar correlacionados. Si lo están, se muestra el mensaje “Correlated”, si no están correlacionados, se da el valor del semi intervalo.

Utilización de los recursos

Para cada recurso, Arena calcula dos estadísticos relacionados con su utilización:

- La *utilización* (“Utilization”), que se define como el promedio temporal de la *utilización instantánea* del recurso.

La *utilización instantánea*, $U(t)$, es el cociente del número de unidades del recurso que están ocupadas en un determinado instante, $B(t)$, entre el número de unidades del recurso que están disponibles en ese mismo instante (libres u ocupadas), $M(t)$:

$$\text{Utilization} = \frac{1}{T} \int_0^T U(t) \cdot dt = \frac{1}{T} \int_0^T \frac{B(t)}{M(t)} \cdot dt \quad (6.1)$$

donde T es la duración de la simulación.

Si el número de recursos disponibles sigue una planificación que intenta adecuar la capacidad a la evolución de su demanda, entonces la utilización es una medida de la bondad de esa adecuación.

- La *utilización planificada* (“Scheduled Utilization”) es el cociente del promedio temporal del número unidades del recurso que están ocupadas entre el promedio temporal del número de unidades del recurso disponibles.

$$\text{NumberBusy} = \frac{1}{T} \int_0^T B(t) \cdot dt \quad (6.2)$$

$$\text{NumberScheduled} = \frac{1}{T} \int_0^T M(t) \cdot dt \quad (6.3)$$

$$\text{ScheduledUtilization} = \frac{\text{NumberBusy}}{\text{NumberScheduled}} \quad (6.4)$$

Esta métrica da una idea de la relación entre ambos y proporciona una medida global de la capacidad total frente a la demanda total, sin considerar la relación temporal entre ambas.

Cuando la capacidad del recurso es fija (M es constante), la *utilización* y la *utilización planificada* son iguales. La aplicación de estas métricas es la siguiente:

- La *utilización planificada* proporciona una idea de en qué medida la capacidad total puede acomodar la demanda.
- La *utilización* mide en qué medida la evolución temporal de la planificación de la capacidad de los recursos sigue a la demanda de capacidad.

Si la *utilización* es mucho más alta que la *utilización planificada*, puede ser una indicación de que se dispone de la capacidad necesaria, pero que su planificación en el tiempo no se ha realizado adecuadamente (no se ajusta a la evolución temporal de la demanda).

Se recomienda al alumno que en este punto resuelva los problemas 6.1, 6.2, 6.3, 6.4 y 6.5

6.3. Modelado detallado

En la sección anterior se ha discutido básicamente el tipo de modelado que puede realizarse empleando los módulos que se encuentran en el panel “Basic Process”. Estos módulos, de “alto nivel”, son relativamente fáciles de usar, con lo cual es una buena estrategia emplearlos en la construcción del modelo hasta donde sea posible, empleando módulos de más “bajo nivel” únicamente en aquellos casos en que sea imprescindible. Arena permite combinar en un mismo modelo módulos de diferentes paneles.

En esta sección se discuten algunos de los módulos de mas “bajo nivel” contenidos en los paneles “Advanced Process” y “Blocks”.

El panel “Advanced Process” complementa al panel “Basic Process”, proporcionando capacidades adicionales que permiten realizar un modelado más detallado. Por ejemplo, la secuencia de los módulos “Seize”, “Delay” y “Release” del panel “Advanced Process” proporciona básicamente la misma lógica de modelado que el módulo “Process” del panel “Basic Process”. La ventaja adicional que proporcionan estos tres módulos del panel “Advanced Process” frente al módulo “Process” es que pueden conectarse formando prácticamente cualquier combinación que requiera el modelo. De hecho, algunos usuarios expertos prefieren iniciar el modelado a este nivel, ya que consideran que el modelo resultante resulta más fácil de entender.

Los paneles “Block” y “Elements” contiene los módulos de más bajo nivel, que corresponden con los bloques del lenguaje de simulación SIMAN, sobre el que se basa Arena. Asimismo, estos dos paneles contienen módulos diseñados para el modelado de sistemas continuos.

El panel “Blocks” proporciona la funcionalidad básica que se ha usado para crear todos los módulos que se encuentran en los tres restantes paneles de Arena: “Basic Process”, “Advanced Process” y “Advanced Transfer”. Además, proporciona construcciones de modelado no disponibles en los demás paneles, tales como los bucles “while”, bifurcaciones en el flujo en las cuales la condición combina la probabilidad y la lógica, etc.

Puede observarse que el nombre de algunos módulos del panel “Blocks” coincide con el nombre de módulos de otros paneles. Aunque los nombres son los mismos (pueden distinguirse por su color y forma), los módulos no son iguales.

Por ejemplo, el módulo “Seize” del panel “Advanced Process” incorpora internamente una cola (lo mismo sucede con el módulo “Process” del panel “Basic Process”), en la que esperan las entidades si el recurso se encuentra ocupado. Sin embargo, el módulo “Seize” del panel “Blocks” no incorpora una cola: es preciso definirla explícitamente, conectando en serie un módulo “Queue” y un módulo “Seize”, ambos del panel “Blocks”.

Otro ejemplo lo constituyen los bloques “Assign” de los paneles “Basic Process” y “Blocks”. Cuando se usa el módulo “Assign” del panel “Basic Process” y se realiza una asignación de un nuevo atributo, Arena define automáticamente el atributo, y lo añade a la lista de atributos del modelo. Por el contrario, el módulo “Assign” del panel “Blocks” no define el atributo, únicamente le asigna un valor.

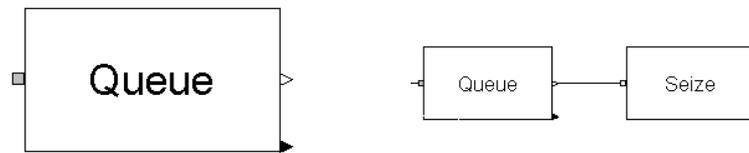


Figura 6.3: Módulo “Queue” del panel “Blocks” (izqda). Conexión “Queue” - “Seize” (drcha).

El panel “Elements” contiene los módulos necesarios para la definición del experimento en SIMAN.

“Balking”

El término inglés “*balking*” se aplica a aquellas situaciones en las cuales se impide el acceso de una entidad a un proceso. A continuación se muestran dos ejemplos.

Ejemplo 6.3.1. *En un sistema de atención telefónica al cliente, si todas las líneas se encuentran ocupadas, el cliente recibe la señal de comunicando, con lo cual debe colgar el teléfono (y posiblemente intentarlo de nuevo más tarde). Se trata de un sistema en el cual no existe cola: si hay un recurso disponible la entidad puede capturarlo, en caso contrario la entidad debe abandonar el proceso.* □

Ejemplo 6.3.2. *Un restaurante de comida rápida tiene un servicio de “drive-through”, atendido desde una sola ventanilla, en el que puede formarse una cola con espacio máximo para 5 vehículos. Mientras haya 5 vehículos en la cola, es decir, mientras ésta tenga su tamaño máximo, no se permite que un sexto vehículo se añada a la cola. Lo que sucede con los vehículos que no pueden acceder al restaurante depende de las hipótesis de modelado: pueden dar una vuelta a la manzana e intentarlo de nuevo, pueden desistir e ir a otro restaurante (es decir, abandonan el sistema), etc.* □

En general, el “*balking*” (traducible como “fracasar”) representa un tipo de fallo del sistema en satisfacer las necesidades de los clientes. Por este motivo, suele ser de interés contar el número de veces que sucede en la simulación. Cuantas menos veces suceda, mejor.

El módulo “Queue” del panel “Blocks” permite definir un tamaño máximo de la cola. Mientras el tamaño de la cola es igual al máximo permitido, las entidades que entran en el módulo “Queue” son inmediatamente obligadas a abandonarlo. El módulo “Queue” del panel “Blocks” posee las dos salidas siguientes (ver el módulo en la parte izquierda de la Figura 6.3):

- Una salida para las entidades que abandonan la cola porque el recurso queda disponible y, por tanto, van a capturarlo (salida situada en el centro del lado derecho del módulo). Esta salida se conecta directamente a la entrada de un módulo “Seize” del panel “Blocks”, que modela la acción de capturar el recurso. La conexión se muestra en la parte derecha de la Figura 6.3.
- Otra salida para las entidades a las que no se permite formar parte de la cola (salida situada en la base del lado derecho del módulo). Esta salida se conecta a los bloques que modelan el flujo de las entidades a las que se impide entrar en la cola.

La forma en que funciona la conexión de los módulos “Queue” - “Seize”, mostrada en la parte derecha de la Figura 6.3, es la siguiente. Cuando una entidad llega al módulo “Queue”, Arena comprueba en primer lugar el estado del recurso del módulo “Seize”. Si el recurso está disponible, entonces no hay entidades esperando en la cola, y la entidad captura el recurso: sale del módulo “Queue” y entra en el módulo “Seize”. Por el contrario, si el recurso no está disponible, entonces la entidad trata de entrar en la cola. Si no hay espacio disponible

en la cola, la entidad sale inmediatamente del módulo “Queue” por la salida que hay en la base del lado derecho de éste.

Se recomienda al alumno que en este punto resuelva el problema 6.6

Conjuntos de objetos

Los *conjuntos* son un medio para referirse a objetos del mismo tipo (recursos, entidades, colas, etc.), que tienen propiedades en común, pero que no son iguales. A continuación se muestra un ejemplo de aplicación de este concepto.

Ejemplo 6.3.3. *Los clientes llaman a un servicio de atención telefónica con el fin de solicitar soporte técnico acerca del producto tipo A o del tipo B. Cinco técnicos atienden este servicio. Los dos primeros están cualificados para responder cuestiones acerca del producto tipo A, el tercero y el cuarto acerca del producto tipo B, y el quinto está cualificado para atender consultas referentes a cualquiera de los dos productos. Puesto que cada cliente accede al servicio con preguntas acerca de un determinado tipo de producto, el cliente debe “capturar” un técnico de entre aquellos que están cualificados para atender cuestiones relativas a este determinado tipo de producto.*

Se definen los dos siguientes conjuntos de recursos:

grupoA: técnico1, técnico2 y técnico5.

grupoB: técnico3, técnico4 y técnico5.

Si se requiere un técnico conocedor del producto A, el recurso debe seleccionarse de entre los miembros del conjunto de recursos grupoA. Igualmente, el recurso deberá seleccionarse del conjunto grupoB si se precisa un técnico cualificado para atender las consultas acerca del tipo de producto B. □

Los módulos

- “Set”, del panel “Basic Process”,
- “Advanced Set”, del panel “Advanced Process”

proporcionan la base para esta funcionalidad. Para definir un conjunto debe proporcionarse la información siguiente:

- El nombre del conjunto.
- El tipo de objetos que componen el conjunto: recurso, contador estadísticos, tipo de entidad, etc.
- Los miembros que componen el conjunto.

La diferencia entre los módulos “Set” y “Advanced Set” está en el tipo de miembros que soportan. Los miembros de un conjunto definido en el módulo “Set” deben ser de uno de los siguientes tipos: recursos, contadores estadísticos (“counters”), estadísticos de tiempo discreto (“tallies”), tipos de entidad o iconos gráficos de entidades. El módulo “Advanced Set” no impone esta restricción: los miembros de un conjunto pueden ser todos ellos de cualquier tipo de objeto de Arena (eso sí, todos ellos del mismo tipo).

Al definir el recurso sobre el que se realiza la acción “Seize”, puede indicarse que el recurso se escoja de entre los miembros de un conjunto. Para ello debe seleccionarse que el tipo de recurso sea “Set”. En este caso, debe también especificarse la regla a seguir cuando, por encontrarse libre, pueda seleccionarse más de un miembro del conjunto. La elección puede hacerse atendiendo a una de varias posibles *reglas de selección*: aleatoriamente, siguiendo un determinado orden de preferencias preestablecido, de forma cíclica, etc.

Se recomienda al alumno que en este punto resuelva el problema 6.7

Variables

Una *variable* representa cierta información que es característica del sistema: no está ligada a una determinada entidad (como sucede con los atributos), sino que pertenece al sistema en su conjunto, y son accesibles desde cualquier parte del sistema. Arena permite que el usuario defina sus propias variables, las cuales pueden tomar valores reales, ser vectores o matrices de dos dimensiones. Esto permite organizar la información en listas o tablas bidimensionales de valores individuales.

Las variables se usan para diferentes propósitos. Por ejemplo, si el tiempo necesario para que una entidad transite entre procesos es el mismo en diferentes partes del sistema, puede asignarse este tiempo a una variable (por ejemplo, llamada *tTrans*) y usarla en todos aquellos módulos en los que se precisa este tiempo. De esta forma, sólo será necesario cambiar el valor de esta variable para modificar el tiempo de tránsito en todo el modelo.

Las variables también pueden representar una magnitud que cambia durante la simulación. Por ejemplo, el número de entidades que se encuentran en determinada zona del sistema, de modo que la variable se incrementa cada vez que una entidad entra en la zona y se decrementa cada vez que una entidad sale de la zona.

Las variables pueden almacenar cantidades reales, y también pueden ser vectores y matrices de dos dimensiones. En el módulo de datos "Variable", del panel "Basic Process", está accesible la información de todas las variables que han sido definidas en el modelo. La definición de las variables puede realizarse:

- En los módulos "Assign" del modelo, con lo que Arena automáticamente registra la información en el módulo "Variables".
- En el módulo "Variable". Este es el procedimiento a seguir para definir variables vectoriales y matriciales, y en aquellos casos en que el valor inicial de la variable deba ser diferente de cero.

La definición de una variable, en el módulo de datos "Variable", requiere introducir la siguiente información:

- El nombre de la variable.
- Su número de filas ("Rows") y columnas ("Columns"). Si se trata de un valor numérico, deben dejarse ambos campos en blanco.
- Su valor inicial. Si no se especifica, la variable toma por defecto el valor inicial cero.

Expresiones

Arena permite al usuario definir *expresiones*, es decir, asociar un nombre a una expresión matemática. Cuando este nombre sea referenciado en el modelo, se evalúa la expresión matemática y se devuelve su valor.

Habitualmente, las expresiones se emplean para obtener valores de una distribución o de una ecuación matemática, en la que pueden intervenir distribuciones, atributos y variables. Cuando una expresión aparece en una sola parte del modelo, es más cómodo introducirla directamente en el módulo donde se precisa. Sin embargo, cuando la expresión se usa en varias partes del modelo, o cuando la forma de la expresión depende del valor de un atributo de una entidad, entonces suele ser más conveniente emplear una expresión.

El módulo "Expressions" de Arena permite definir las expresiones y sus valores asociados. Al igual que las variables, las expresiones son referenciadas por su nombre, y pueden ser especificadas como vectores y matrices bidimensionales.

Submodelos

Cuando se desarrollan modelos complejos, a menudo es útil dividir el modelo en modelos más pequeños, denominados *submodelos*, que pueden o no interactuar entre sí. Es decir, los submodelos pueden conectarse entre sí, o bien pueden permanecer aislados dentro del modelo.

Además de la capacidad de *modularidad*, Arena permite estructurar el modelo de manera *jerárquica*: los submodelos pueden, a su vez, contener otros submodelos y así sucesivamente. Arena facilita la navegación entre los diferentes submodelos: en la sección “Navigate” de la barra del proyecto se muestra, en forma de árbol, la estructura completa de los submodelos.

Este tipo de estrategia de modelado, denominada *modelado modular y jerárquico*, facilita la realización, la comprensión, el depurado y el mantenimiento de los modelos.

Para crear un submodelo, debe seleccionarse la opción *Object / Submodel / Add Submodel*. Entonces, el cursor adquiere la forma de cruz. Haciendo “clic” con el botón izquierdo del ratón sobre la ventana del modelo, se sitúa el submodelo. Haciendo “clic” con el botón derecho sobre el icono del submodelo se despliega un menú, en el que puede seleccionarse:

- *Editar las propiedades* del submodelo, como son su nombre, el número de puntos de entrada y el número de puntos de salida. Si el submodelo no interactúa con el resto del modelo, el número de puntos de entrada y de salida es cero.
- *Abrir el submodelo*, con lo que se desciende un nivel en la jerarquía del modelo. En la pantalla se muestran los puntos de entrada y de salida del submodelo, y puede definirse el diagrama de módulos que representa el comportamiento interno del submodelo.

Costes

Arena calcula automáticamente el tiempo y el coste acumulado por cada entidad en su tránsito por el sistema, distinguiendo entre los costes debidos a cada uno de los cinco conceptos en los que se puede asignar el tiempo: “Value Added”, “Non-Value Added”, “Transfer”, “Wait” y “Other”.

Para obtener estadísticas del coste válidas, es preciso introducir la información de los costes en el modelo. La información básica del coste se introduce en los módulos de datos “Entity” y “Resource”.

EJERCICIOS

6.1 Realice el estudio de simulación descrito a continuación², usando para ello el entorno de modelado Arena.

El modelo representa un paso en una cadena de fabricación, consistente en hacer un agujero en las piezas mediante una única máquina perforadora. Las piezas llegan a este paso, son procesadas por un único recurso y abandonan el proceso. Si la perforadora se encuentra libre en el instante en que llega la pieza, entonces ésta es procesada inmediatamente. En caso contrario, la pieza espera su turno en una cola con disciplina FIFO.

El intervalo de tiempo entre llegadas sucesivas de piezas está distribuido exponencialmente, con media 5 minutos. Las piezas llegan de una en una. El tiempo de proceso está distribuido triangularmente, con rango de 1 a 6 minutos y modo 3 minutos.

Condiciones iniciales: la máquina está libre y la cola vacía. Condición de finalización: el tiempo simulado alcanza el valor 20 minutos. El objetivo del estudio es estimar los estadísticos siguientes:

- Número total de piezas procesadas.

²Este estudio de simulación está extraído del texto (Kelton et al. 2002).

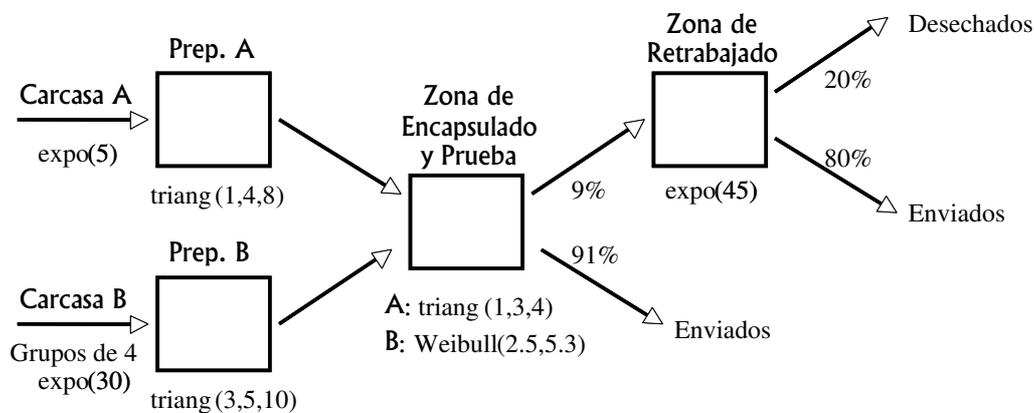


Figura 6.4: Sistema de ensamblado y prueba de circuitos electrónicos.

- Tiempo promedio de espera en la cola.
- Tiempo máximo de espera en cola.
- Número medio de piezas esperando en la cola.
- Tamaño máximo de la cola.
- Tiempo de ciclo medio, es decir, el valor esperado del tiempo que transcurre desde que la pieza llega al sistema hasta que lo abandona.
- Tiempo de ciclo máximo.
- Utilización de la máquina perforadora, es decir, proporción del tiempo que se encuentra ocupada.

6.2 Realizar el estudio de simulación descrito a continuación³, usando para ello el entorno de modelado Arena.

Parte A. El modelo representa el final del proceso de fabricación de un circuito electrónico: su montaje en el interior de una carcasa metálica y la realización de pruebas eléctricas para determinar si el dispositivo, una vez encapsulado, funciona correctamente. Si bien se trata de un único tipo de dispositivo electrónico, éste puede encapsularse de dos maneras: usando la carcasa “Tipo A” o usando la carcasa “Tipo B”.

El flujo del modelo, representado en la Figura 6.4, es el siguiente:

- Las carcasas metálicas, ya preparadas para alojar el circuito en su interior, llegan al sistema. Las carcasas de Tipo A llegan de una en una, mientras que las de Tipo B llegan en grupos de 4. Los intervalos de tiempo entre llegadas sucesivas están distribuidos exponencialmente, con media 5 minutos y 30 minutos respectivamente.
- Al llegar al sistema, las carcasas deben recibir cierto tratamiento (pulido de los bordes, limpieza, etc.) en la *Zona de Preparación*. El tratamiento y el recurso que lo realiza dependen del tipo de carcasa:
 - En la *Zona Prep. A* se preparan las carcasas de Tipo A. El tiempo de proceso está distribuido triangularmente, con rango entre 1 y 8 minutos y modo igual a 4 minutos.
 - En la *Zona Prep. B* se preparan las carcasas de Tipo B. El tiempo de proceso está distribuido triangularmente, con rango entre 3 y 10 minutos, y modo 5 minutos

En ambos casos, las carcasas son procesadas una a una. En particular, las carcasas de Tipo B, que son recibidas en grupos de cuatro, son separadas y procesadas individualmente.

³Este estudio de simulación está extraído del texto (Kelton et al. 2002).

- Una vez concluido el proceso de preparación, las carcasas de ambos tipos van a la *Zona de Encapsulado y Prueba*. El proceso en esta zona consiste en introducir el circuito dentro de la carcasa, cerrar y sellar ésta y probar eléctricamente el dispositivo. La distribución de probabilidad del tiempo empleado en este proceso depende del tipo de carcasa. Para las de Tipo A está distribuido triangularmente, con rango de 1 a 4 minutos y con 3 minutos de modo. EL tiempo de proceso de los dispositivos con carcasa Tipo B está distribuido Weibull, con $\alpha = 2.5$ minutos y $\beta = 5.3$ minutos.
- EL 91% de los dispositivos pasa las pruebas eléctricas y son enviados. Se asume que la probabilidad de fallo de un dispositivo es independiente de probabilidad de fallo de los demás dispositivos. Los dispositivos fallados son enviados a la *Zona de Retrabajado*.
- En la *Zona de Retrabajado* los circuitos son extraídos de las cajas, reparados, vueltos a encapsular y probados de nuevo. El 80% de los dispositivos re trabajados pasan con éxito este nuevo test y son enviados. El 20% restante no consigue ser reparado y es desechado. Se considera que el tiempo del proceso del re trabajado es independiente del tipo de carcasa y de si finalmente se consigue reparar el dispositivo o no. El tiempo de re trabajado está distribuido exponencialmente, con media 45 minutos.

La cadena de encapsulado opera durante 2 turnos al día, de 8 horas cada uno. Se considera que el funcionamiento en ambos turnos es similar, con lo cual el modelo no depende del turno. Asimismo, puesto que las condiciones al comienzo de un turno son las mismas que al finalizar el turno anterior, puede realizarse la simulación sin solución de continuidad entre turnos.

Condiciones iniciales: todas las colas están vacías y todos los recursos libres. Condición de finalización: la duración de la simulación será 4 turnos, de 8 horas/turno (es decir, 1920 minutos).

El objetivo del estudio es estimar los estadísticos siguientes:

- La utilización de los recursos.
- El tamaño medio de cada cola.
- El tiempo medio en cada cola.
- Los tiempos de ciclo de los dispositivos enviados sin re trabajar, de los re trabajados y de los dispositivos desechados.

Parte B. La cadena de encapsulado de la Parte A del problema opera durante 2 turnos al día, de 8 horas cada uno, que eran totalmente equivalentes entre sí. Ahora se supone que la operación de la Zona de Retrabajado cambia de un turno a otro: en el primer turno trabaja un solo operario, mientras que en el segundo turno trabajan dos. Repetid el estudio de simulación del problema anterior, introduciendo esta modificación en el modelo y ampliando el periodo de estudio a 10 días. El objetivo es estimar:

- El tiempo medio en la cola del recurso de re trabajado.
- Los tiempos de ciclo de los dispositivos enviados sin re trabajar, de los re trabajados y de los dispositivos desechados.

Parte C. De cuando en cuando el recurso de la Zona de Encapsulado y Prueba se estropea. De los datos recogidos, se hace la hipótesis de que el intervalo de tiempo entre fallos consecutivos está distribuido exponencialmente, con media 120 minutos. El tiempo requerido para arreglar la avería es también una variable aleatoria, distribuida exponencialmente con media 4 minutos. Ampliad el modelo de simulación de la Parte B, de modo que contemple este tipo de averías, y empleadlo para realizar el estudio descrito a continuación. El estudio tiene los objetivos siguientes:

- Obtener información acerca del tiempo que el recurso de la Zona de Encapsulado y Prueba pasa fuera de servicio.
- Se planea comprar estanterías para almacenar ordenadamente los dispositivos que están en cola en la Zona de Retrabajado. La capacidad de cada estantería es de 10 dispositivos. Se pretende determinar cuántas de ellas es preciso comprar. Para ello, se desea estimar mediante simulación durante cuánto tiempo el número de dispositivos en cola es cero, durante cuánto tiempo es mayor que cero y menor o

igual que 10, durante cuánto tiempo es mayor que 11 y menor o igual que 20, y así sucesivamente.

- Obtener información acerca de la utilización de los recursos del modelo.

6.3 Realizar el estudio de simulación descrito a continuación⁴, usando para ello el entorno de modelado Arena.

Parte A. Los pasajeros llegan a la puerta principal de la terminal de un aeropuerto, y a continuación van al mostrador de facturación. Una vez finalizada la facturación se dirigen a sus puertas de embarque.

- El intervalo de tiempo entre llegadas sucesivas de pasajeros a la puerta principal está distribuido exponencialmente, con media 1.6 minutos. Los pasajeros llegan de uno en uno.
- El tiempo que tardan los pasajeros en ir de la puerta principal al mostrador de facturación está distribuido uniformemente entre 2 y 3 minutos.
- En el mostrador de facturación esperan en una cola FIFO hasta que uno de los 5 empleados quede disponible para atenderles.
- El tiempo del proceso de facturación está distribuido (en minutos) Weibull con parámetros $\beta = 7.76$ y $\alpha = 3.91$.

Condiciones iniciales: la cola está vacía y los recursos libres. Condición de finalización: la duración de la simulación será de 16 horas, que es el tiempo diario de funcionamiento del mostrador de facturación. El objetivo del estudio es estimar los estadísticos siguientes:

- Tiempo promedio de los pasajeros en el sistema.
- Número de pasajeros que han completado la facturación durante el tiempo simulado.
- Número medio de pasajeros que esperan en la cola del mostrador de facturación.

Parte B. Los empleados trabajan en turnos de 8 horas. Realizan descansos de 15 minutos de manera escalonada, empezando a los 90 minutos de entrar en el turno. El descanso para la comida dura 30 minutos y también se hace de forma escalonada, empezando a las 3 horas y media de haber comenzado el turno. Repetir el estudio considerando esta modificación en el modelo.

Parte C. Repetir el estudio considerando la siguiente modificación en el modelo. Existen dos tipos diferentes de pasajeros:

- El primer tipo llega a la puerta principal de acuerdo a una distribución exponencial de media 2.4 minutos y su tiempo de facturación sigue una distribución gamma con parámetros $\beta = 0.42$ y $\alpha = 14.4$ (expresado en minutos).
- El segundo tipo de pasajero llega con una distribución exponencial de media 4.4 minutos y su tiempo de facturación es igual a $3 + X$ minutos, donde X está distribuido Erlang con parámetros $\text{ExpMean} = 0.54$ y $k = 15$.

6.4 Realizar el estudio de simulación descrito a continuación⁵, usando para ello el entorno de modelado Arena.

Dos tipos diferentes de piezas llegan a un sistema para ser procesadas por un mismo operario. Cada tipo de cola espera en su propia cola, hasta que el único operario del sistema queda libre. Una vez procesadas, ambos tipos de piezas son sometidas a un segundo proceso, realizado por una máquina, concluido el cual, las piezas abandonan el sistema. Las distribuciones de probabilidad de los intervalos de tiempo entre llegadas, y de la duración de la fase "Delay" de los procesos, son las siguientes (todos los tiempos están expresados en horas):

⁴Este estudio de simulación está extraído del texto (Kelton et al. 2002).

⁵Este estudio de simulación está extraído del texto (Kelton et al. 2002).

| | Proceso de llegada | Primer proceso | Segundo proceso |
|----------|--|---|---|
| Piezas A | Lognormal(LogMean=11.5, LogStd=2.0) | Triangular(Min=5, Mode=6, Max=8) | Triangular(Min=4, Mode=6, Max=8) |
| Piezas B | Exponencial(Mean=15) | Triangular(Min=3, Mode=7, Max=8) | Triangular(Min=4, Mode=6, Max=8) |

Condiciones iniciales: no hay piezas en el sistema (colas vacías y recursos libres). Condición de finalización: la duración de la simulación será de 5000 horas. El objetivo del estudio es estimar los estadísticos siguientes:

- Tiempo de ciclo medio de las piezas.
- Número medio de piezas esperando en las dos colas del primer proceso.

6.5 Realizar el estudio de simulación descrito a continuación⁶, usando para ello el entorno de modelado Arena.

Se desea modelar el funcionamiento de un restaurante de comida rápida, en el periodo de tiempo que va desde las 10 AM hasta las 2 PM, que es en el que se produce una mayor afluencia de clientes. La llegada de clientes se produce de las tres maneras siguientes:

- *Andando*. Los clientes llegan de uno en uno, con intervalos entre llegadas distribuidos exponencialmente, con media 3 minutos. La primera llegada se produce, a partir de las 10 AM, en un instante aleatorio distribuido EXPO(3 minutos).
- *En coche*. El número de clientes por coche puede ser 1, 2, 3 ó 4, con probabilidades 0.2, 0.3, 0.3 y 0.2 respectivamente. Los intervalos entre llegadas están distribuidos exponencialmente con media 5 minutos. La primera llegada se produce, a partir de las 10 AM, en un instante aleatorio distribuido EXPO(5 minutos).
- *En autobús*. Cada día llega un único autobús, en un instante aleatorio distribuido uniformemente entre las 11 AM y las 1 PM. El número de pasajeros del autobús varía de un día a otro, y al parecer sigue aproximadamente una distribución de Poisson con media 30 pasajeros.

Una vez que los clientes llegan al restaurante, con independencia del medio empleado para ello, operan independientemente. En primer lugar se dirigen al mostrador de *petición/pago*, en cual tardan TRIA(1,2,4) minutos en realizar su petición y TRIA(1,2,3) minutos en pagarla. Las operaciones de petición y de pago se realizan secuencialmente: en primer lugar la petición, y a continuación el pago, en las cuales el cliente es atendido por un mismo empleado.

A continuación, el cliente se dirige al mostrador de *recogida*, con el fin de recoger la comida que han pedido. El tiempo que transcurre entre que comienza a ser atendido en dicho mostrador y el instante en que se le entrega la comida está distribuido uniformemente entre 30 segundos y 2 minutos.

Seguidamente, el cliente se dirige al comedor, en el cual hay 30 asientos. Cuando queda un asiento libre, el cliente se sienta en él (no necesariamente con las personas de su grupo). Una vez sentado, comienza a comer, empleando un tiempo TRIA(10,20,30) minutos en terminar su comida, hecho lo cual abandona el restaurante.

En cada mostrador existe una cola FIFO: en el mostrador de *petición/pago* y en el mostrador de *recogida*. Asimismo, hay otra cola FIFO en la que los clientes esperan a que quede libre un asiento en el comedor.

El tiempo de tránsito de la puerta al mostrador de *petición/pago* está distribuido EXPO(30) segundos. Igualmente, está distribuido EXPO(30) segundos el tiempo para ir de este primer mostrador al de *recogida*, y el tiempo necesario para ir de este segundo mostrador al comedor. Después de comer, el cliente se mueve algo más lentamente, con lo cual tarda EXPO(1) minuto en ir del comedor a la puerta de salida del restaurante.

⁶Este estudio de simulación está extraído del texto (Kelton et al. 2002).

Durante las 4 horas que dura el periodo de tiempo bajo estudio, 6 empleados atienden el mostrador de *petición/pago* y otros 2 empleados el mostrador de *recogida*.

Condiciones iniciales: inicialmente la máquina está libre y la cola vacía. Condición de finalización: se desea simular el sistema durante las 4 horas que dura el periodo bajo estudio. El objetivo del estudio es estimar los siguientes indicadores de la congestión del sistema:

- Longitud promedio y máxima de cada cola.
- Tiempo de espera promedio y máximo en cada cola.
- Número total de clientes que abandonan el restaurante.

- 6.6** Realizar el estudio de simulación descrito a continuación, usando para ello el entorno de modelado Arena. El sistema a estudiar es un servicio de atención telefónica completamente automatizado, cuya función es informar al cliente acerca de tres tipos de producto: A, B y C.

El cliente llama a un número central, que es atendido por 10 líneas telefónicas. Si en ese instante todas las líneas se encuentran ocupadas, el cliente debe colgar el teléfono, es decir, abandonar el sistema. Si alguna de las líneas telefónicas se encuentra libre, el cliente la ocupa y escucha una grabación, en la que se le ofrece la opción de recibir información sobre uno de los tres tipos de producto: A, B y C. Los clientes escogen entre estas tres opciones con probabilidad 0.76, 0.16 y 0.08 respectivamente. El tiempo empleado en esta actividad está distribuido uniformemente entre 0.1 y 0.6 minutos.

A continuación, el cliente escucha la grabación correspondiente al tipo de producto que ha escogido. Durante la grabación, se le pide que introduzca determinados datos, empleando el teclado telefónico. El tiempo que tarda el cliente en completar la audición de la grabación depende del tipo de producto, y de las elecciones que ha ido haciendo a través del teclado. En los tres productos, el tiempo que dura la audición está distribuido triangularmente, con los siguientes valores del rango y el modo (expresados en minutos):

| Producto | Rango | Modo |
|----------|---------|------|
| A | 10 - 18 | 15 |
| B | 8 - 20 | 15 |
| C | 7 - 12 | 10 |

La sistema de atención telefónica funciona durante 8 horas al día. La *frecuencia de llegada de llamadas* se supone constante. El tiempo entre llamadas está distribuido exponencialmente, con media 1 minuto. El tiempo que transcurre entre la apertura del sistema y la primera llamada también está distribuido exponencialmente, con media 1 minuto.

Condiciones iniciales: al comenzar el día, no hay ninguna llamada en el sistema. Condición de finalización: desea simularse el sistema durante las 8 horas que permanece en funcionamiento cada día. Objetivo del estudio: como medida del grado de satisfacción de los clientes con el servicio, se desea estimar el número de clientes que han encontrado todas las líneas ocupadas y que han tenido que colgar sin ser atendidos.

- 6.7** Realizar el estudio de simulación descrito a continuación⁷, usando para ello el entorno de modelado Arena.

El sistema bajo estudio es un servicio telefónico que presta soporte técnico sobre tres tipos de producto (1, 2 y 3), y en el que trabajan 11 técnicos. El cliente accede a este servicio llamando a un número telefónico central, que es atendido por 26 líneas. Si un cliente llama en un instante en que las 26 líneas se encuentran ocupadas, entonces recibe la señal de comunicando y debe colgar el teléfono. Si, por el contrario, alguna línea se encuentra libre, entonces el cliente la ocupa y escucha una grabación, en la que se le pregunta sobre qué producto desea realizar la consulta: 1, 2 ó 3.

Los porcentajes de consultas acerca de cada uno de los tres tipos de producto son el 25%, 34% y 41% respectivamente. Esta actividad requiere un tiempo distribuido uniformemente entre 0.1 y 0.5 minutos.

⁷Este estudio de simulación está extraído del texto (Kelton et al. 2002).

No todos los técnicos están cualificados para responder preguntas sobre los tres tipos de producto. Si un técnico cualificado para prestar soporte técnico de ese tipo de producto se encuentra libre, entonces el cliente es atendido. Si ninguno se encuentra libre, el cliente entra en una cola electrónica, en la que escucha música hasta que un técnico capacitado para atenderle queda libre.

El tiempo necesario para contestar una pregunta técnica está distribuido triangularmente, con rango de 3 a 18 minutos y modo 6 minutos, con independencia del tipo de producto al que haga referencia. Una vez completa la consulta, el cliente abandona el sistema.

La sistema de atención telefónica funciona desde las 8 AM hasta las 7 PM. Aunque el sistema cierra a las 7 PM, todas las llamadas que hasta ese instante han entrado son atendidas. El tiempo entre llamadas sucesivas está distribuido exponencialmente, con media 1 minuto. El tiempo que transcurre desde las 8 AM hasta que se recibe la primera llamada también está distribuido exponencialmente con media 1 minuto.

Cada uno de los 11 técnicos trabaja 8 horas al día, y dispone de media hora para comer (no incluida en las 8 horas). No todos ellos están cualificados para responder cuestiones acerca de los tres tipos de producto. En la siguiente tabla se muestra el tipo de llamadas que puede atender cada uno de ellos, así como su horario.

| Técnico núm. | Tipo de producto | Periodo de tiempo (30 minutos) | | | | | | | | | | | | | | | | | | | | | |
|-----------------|---------------------|--------------------------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 1 | 1 | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | | | | | |
| 2 | 1 | | | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 3 | 1,3 | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 4 | 1,2,3 | | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 5 | 1,2,3 | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 6 | 2 | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 7 | 2 | | | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 8 | 2 | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 9 | 3 | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 10 | 3 | | | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |
| 11 | 3 | | | | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • | • |

Condiciones iniciales: no hay ninguna entidad en el sistema. Condición de finalización: que la simulación alcance 11 horas, que es el periodo de funcionamiento diario del servicio (desde las 8 AM hasta las 7 PM).

Interesa estimar los siguientes estadísticos, que miden el grado de satisfacción de los clientes con el servicio:

- El número de clientes que han encontrado todas las líneas ocupadas y que han tenido que colgar sin ser atendidos.
- En función del tipo de producto acerca del que se realiza la consulta:
 - El tiempo medio que debe esperar el cliente (escuchando música) para recibir la atención de un técnico.
 - Número medio de llamadas en espera de ser atendidas.

Asimismo, se desea conocer la utilización de cada una de los técnicos, a fin de evaluar su carga de trabajo.

Tema 7

Modelos analíticos y simulación

NOTA IMPORTANTE: Este tema NO SE EXIGIRÁ EN EL EXAMEN durante el presente curso académico, si bien se recomienda su lectura. Por este motivo, no se plantean objetivos docentes del tema, ni tampoco se proponen problemas.

7.1. Introducción

Como se ha indicado en los temas anteriores, los modelos matemáticos constituyen una abstracción del sistema real. Las expresiones matemáticas se emplean para describir aquellas relaciones entre los elementos del sistema que el ingeniero considera relevantes para el objetivo del estudio. Un modelo formulado adecuadamente debe reflejar aquellas preguntas que el ingeniero desea responder acerca del sistema real. Si el modelo puede ser usado para responder estas cuestiones mediante la aplicación de métodos analíticos de matemáticas (tales como álgebra o cálculo), entonces se dice que el modelo es *analítico*. Por el contrario, si sólo pueden obtenerse las respuestas, a partir del modelo, empleando métodos de análisis numérico, entonces se dice que el modelo es *numérico* o de *simulación*.

En los estudios de simulación se usan normalmente ambos tipos de modelos, ya que proporcionan diferentes perspectivas del sistema bajo estudio. Los modelos analíticos proporcionan información (de manera relativamente rápida) acerca del comportamiento del sistema, y son útiles en la construcción, validación, verificación y análisis de los modelos de simulación. Por consiguiente, el empleo de modelos analíticos no debe ser considerado como una alternativa al empleo del modelo de simulación, sino como una parte integral del estudio de simulación

En esta sección:

- se discuten brevemente las ventajas y desventajas de los modelos analíticos y numéricos,
- se presenta una clase de modelos denominada *Cadenas de Markov*, y
- se exponen las propiedades de dos tipos de modelos de tiempo discreto: los modelos de colas y los modelos de inventario.

Aunque se presentarán algunas de las propiedades más importantes de los modelos analíticos, no se incluye la demostración formal de las mismas. Puede encontrarse un tratamiento más completo y detallado de los modelos analíticos en los textos clasificados en el campo “Investigación Operativa” (“Operations Research” en lengua inglesa).

7.2. Modelos analíticos frente a modelos de simulación

A continuación se discuten algunas de las ventajas y desventajas de los modelos analíticos y de simulación.

Ventajas y desventajas de los modelos analíticos

Una dificultad importante de los modelos analíticos es que para su construcción y comprensión se requieren unos conocimientos considerables de matemáticas. Existen, no obstante, modelos matemáticos muy bien documentados aplicables a sistemas de colas, sistemas de estado finito, sistemas de inventario, etc.

Los modelos analíticos se emplean en la verificación de los modelos de simulación. A menudo los parámetros y las hipótesis de los modelos de simulación pueden adaptarse de modo que correspondan con los de un modelo analítico sencillo. Bajo estas circunstancias, puede juzgarse si la lógica del modelo de simulación es correcta, comparando las salidas del modelo de simulación con los resultados del modelo analítico.

Asimismo, los modelos analíticos pueden usarse para determinar en qué dirección cambia el comportamiento del sistema debido a un determinado cambio en sus parámetros. Si el modelo de simulación no muestra el mismo comportamiento general que el modelo analítico, existen razones para sospechar que se ha cometido un error en el modelo de simulación: deben investigarse cuáles son las razones de este comportamiento inesperado. Aun en el caso en el cual finalmente se encontrara una explicación razonable al comportamiento del modelo de simulación, el modelo analítico habría contribuido a la validación del modelo de simulación.

En algunos problemas, los modelos analíticos permiten obtener un tipo de respuestas diferentes de las que permiten obtener los modelos de simulación. Para ilustrar esta afirmación, se va a emplear un problema denominado la *Paradoja de San Petersburgo*¹.

Ejemplo 7.2.1. *La paradoja se plantea en un juego de azar, en el cual se lanza una moneda hasta que aparece por primera vez “cara”. El dinero que se paga al participante es 2^x euros, donde x es el número de lanzamientos realizado hasta que obtuvo “cara” por primera vez (incluyendo el lanzamiento en que obtuvo cara).*

La cuestión que se plantea es: ¿cuál es el precio justo que debe pagar el jugador para poder participar en el juego?

El precio debe ser al menos igual a 2 euros, ya que es el mínimo premio que recibe el participante (en caso de obtener cara en el primer lanzamiento).

Para determinar el precio justo, se construye un modelo analítico del juego, que permita calcular el valor esperado del dinero que recibe el jugador. Llamando x al número de lanzamientos de la moneda hasta que se obtiene por primera vez “cara”, entonces:

$$\text{Prob}\{x\} = (\text{Prob}\{\text{cruz}\})^{x-1} \cdot \text{Prob}\{\text{cara}\} \quad (7.1)$$

A partir de esta expresión, puede calcularse el valor esperado del dinero que percibe el jugador:

¹El ejemplo ha sido extraído del texto (Hoover & Perry 1989).

$$\begin{aligned}
 E\{2^x\} &= \sum_{x:1}^{\infty} 2^x \cdot (\text{Prob}\{\text{cruz}\})^{x-1} \cdot \text{Prob}\{\text{cara}\} & (7.2) \\
 &= \sum_{x:1}^{\infty} 2^x \cdot \left(\frac{1}{2}\right)^{x-1} \cdot \frac{1}{2} \\
 &= \sum_{x:1}^{\infty} 1 \\
 &= \infty
 \end{aligned}$$

Aquí está la paradoja: aunque aproximadamente la mitad de los jugadores perciben solamente 2 euros, el valor esperado del dinero que recibe un jugador no está acotado. En otras palabras, el modelo analítico indica que no existe un precio suficientemente alto para poder jugar.

Un argumento en contra de este análisis es que haría falta estar jugando eternamente para que el jugador percibiera una ganancia infinita. Sin embargo, éste es un argumento débil, ya que puede modificarse el modelo de modo que se considere un número máximo de tiradas de, por ejemplo, un millón:

$$\text{Dinero que percibe el jugador} = \begin{cases} 2^x & \text{si } x = 1, 2, \dots, 10^6 \\ 0 & \text{si } x > 10^6 \end{cases} \quad (7.3)$$

Y aun en este caso, el valor esperado del dinero que recibe el jugador es igual a 1 millón de euros. ¿Pagaría usted la mitad de ese dinero por participar en el juego? □

El modelo analítico proporciona una visión valiosa del juego, pero no proporciona una visión completa.

Ventajas y desventajas de los modelos de simulación

En los modelos de simulación, la relación entre los elementos del sistema se expresa mediante algoritmos codificados en un programa de ordenador. Para determinar el estado del sistema en algún instante de tiempo futuro, se emplea el programa de ordenador para ejecutar el algoritmo. Los modelos de simulación no proporcionan respuestas directas, pero revelan las propiedades del sistema al ejecutar el programa de ordenador. Si el algoritmo incluye la naturaleza estocástica del sistema, entonces la salida del programa debe ser analizada estadísticamente antes de poder extraer conclusiones acerca de las propiedades del sistema.

Una de las grandes ventajas de los modelos de simulación es su flexibilidad. Dado que los modelos de simulación están codificados en programas de ordenador, pueden incluir muchos de los elementos y características del sistema que no están incluidos en los modelos analíticos. En los modelos de simulación, las propiedades del sistema no son determinadas mediante análisis matemático directo, pudiéndose por tanto incluir en el modelo elementos y relaciones que harían el modelo analítico matemáticamente intratable.

Una ventaja adicional de los modelos de simulación es su naturaleza algorítmica. Normalmente es más fácil para los no expertos en matemáticas entender el funcionamiento del modelo cuando éste es un algoritmo que cuando es un sistema de ecuaciones. Esto es importante a la hora de validar el modelo y también a la hora de implementar las conclusiones del estudio: en la medida en que el modelo sea más fácilmente comprensible, mayor será su credibilidad.

Sin embargo, modelar un sistema mediante un modelo de simulación tiene sus desventajas. La construcción y validación de los modelos de simulación no suele ser una tarea ni

rápida ni barata. Puede transcurrir un tiempo considerable desde que comienza a diseñarse el modelo de simulación hasta su construcción y validación. El análisis de las salidas del modelo de simulación es una tarea difícil cuando el modelo incluye varios elementos estocásticos: la simulación proporciona únicamente una estimación de las propiedades del sistema, precisándose el uso de técnicas estadísticas para extraer conclusiones.

A continuación, se discute una aproximación alternativa al problema que conduce a la Paradoja de San Petersburgo.

Ejemplo 7.2.2. *Se aborda el problema construyendo un modelo de simulación del juego, y simulando múltiples veces el desarrollo del mismo. En cada réplica de la simulación se guarda registro del dinero pagado a cada jugador. Los resultados obtenidos son los siguientes:*

| <i>Número de jugadas simuladas</i> | <i>Promedio pagado por jugador (euros)</i> | <i>Máximo pagado en una jugada (euros)</i> |
|------------------------------------|--|--|
| 10 | 5.60 | 32 |
| 50 | 4.90 | 1024 |
| 100 | 13.56 | 256 |
| 500 | 21.48 | 4096 |
| 1000 | 11.94 | 1024 |
| 10000 | 12.56 | 8192 |

La simulación aporta nueva luz sobre el juego. Aunque las ganancias potenciales son muy elevadas, la probabilidad de realizar grandes ganancias es tan pequeña que aun después de haber jugado 10000 jugadores, el promedio del dinero que ha percibido cada uno está apenas por encima de 12 euros. □

El empleo de ambos modelos (analítico y de simulación) ha contribuido a proporcionar una mejor comprensión del juego. Ahora se dispone de un criterio para establecer el precio que debe pagar cada jugador por el derecho a jugar una partida. Aunque la ganancia potencial de un jugador es ilimitada, un jugador juicioso no pagaría mucho más de 12 euros por jugar una partida.

7.3. Modelos analíticos de cambio de estado: cadenas de Markov

Existe gran cantidad de sistemas que pueden modelarse como sistemas de eventos discretos que realizan transiciones entre un conjunto finito de estados. Algunos ejemplos son los siguientes:

- *Sistema: un ascensor. Estado:* el número de piso en que se encuentra.
- *Sistema: el inventario de un almacén. Estado:* número de unidades de producto que están almacenadas.
- *Sistema: red de ordenadores. Estado:* número de procesos activos.
- *Sistema: un hospital. Estado:* número de camas ocupadas.

El estado de un sistema puede estar definido por más de una variable. Cuando está definido por más de 2 variables es complicado desarrollar modelos analíticos del sistema. No obstante, si el estado del sistema es unidimensional y el sistema tiene la *propiedad de Markov*, entonces normalmente es sencillo construir modelos analíticos, que resultan muy útiles para comprender el comportamiento del sistema.

Para definir qué se entiende por un *sistema de Markov*, es conveniente introducir previamente la notación siguiente:

| | |
|----------------------------------|---|
| E_i , con $i : 1, 2, \dots$ | Conjunto de estados en los que puede encontrarse el sistema. |
| t_n , con $n : 0, 1, 2, \dots$ | Instantes en los que el sistema realiza las transiciones de estado. |
| $X(t_n)$ | Estado del sistema en el instante t_n . |

Se dice que el sistema es de Markov si satisface la condición siguiente:

$$\text{Prob}\{X(t_{n+1}) = E_i | X(t_n), X(t_{n-1}), \dots, X(t_0)\} = \text{Prob}\{X(t_{n+1}) = E_i | X(t_n)\} \quad (7.4)$$

La expresión anterior equivale a afirmar que la probabilidad de que el sistema realice la transición al estado E_i únicamente depende del estado actual del sistema, y no de los estados anteriores. Se dice que este tipo de sistemas “no tienen memoria”, ya que el estado futuro del sistema no depende del camino que se siguió hasta llegar al estado actual, sólo depende del estado actual.

En los sistemas de Markov, la probabilidad de que se produzca una transición del estado E_i al estado E_j viene dada por la expresión siguiente:

$$p_{i,j} = \text{Prob}\{X(t_{n+1}) = E_j | X(t_n) = E_i\} \quad (7.5)$$

Si el número de estados, n , es finito, entonces puede definirse la *matriz de transiciones* de la forma siguiente:

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix} \quad (7.6)$$

La probabilidad de que el sistema ocupe cada uno de sus n estados, después de haber realizado k transiciones, se representa mediante el siguiente vector de probabilidades:

$$\pi^k = (\text{Prob}\{X(t_k) = E_1\}, \text{Prob}\{X(t_k) = E_2\}, \dots, \text{Prob}\{X(t_k) = E_n\}) \quad (7.7)$$

Puede demostrarse que:

$$\pi^{k+1} = \pi^k \cdot \mathbf{P} \quad (7.8)$$

con lo cual:

$$\pi^k = \pi^0 \cdot \mathbf{P}^k \quad (7.9)$$

donde \mathbf{P}^k es el producto de la matriz \mathbf{P} por sí misma k veces.

El elemento i, j de la matriz \mathbf{P}^k es:

$$p_{i,j}^k = \text{Prob}\{X(t_k) = E_j | X(t_0) = E_i\} \quad (7.10)$$

Representa la probabilidad en el instante t_k el sistema esté en el estado E_j si en el instante t_0 el sistema se encontraba en el estado E_i .

A continuación se muestran dos ejemplos de sistemas de Markov:

- El modelo de un concesionario de compra y venta de vehículos usados.
- El modelo de la cuota de mercado de un producto.

Ejemplo 7.3.1. *Considérese un pequeño concesionario de compra y venta de coches de segunda mano, con capacidad para albergar, como máximo, 2 coches.*

Las ventas de coches diarias a los clientes se producen con las siguientes probabilidades:

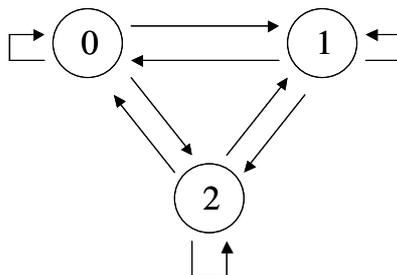


Figura 7.1: Diagrama de transición.

- La probabilidad de que a lo largo del día acuda únicamente 1 cliente para comprar un coche es igual a 0.2.
- La probabilidad de que en un día acudan 2 clientes para comprar sendos coches es igual a 0.04.
- La probabilidad de que en un mismo día se presenten más de 2 compradores es despreciable.
- La probabilidad de no realizar ninguna venta en todo el día es igual a $(1 - 0.2 - 0.04) = 0.76$.

Asimismo, las compras de coches a los clientes en un día obedecen a las siguientes probabilidades:

- La probabilidad de que en todo el día acuda 1 cliente para vender un coche es 0.3.
- La probabilidad de que acudan 2 clientes es 0.09.
- La probabilidad de que acudan más de dos vendedores en un mismo día es despreciable.
- La probabilidad de que en un día no acuda ningún cliente dispuesto a vender su coche es igual a $(1 - 0.3 - 0.09) = 0.61$.

Además, el funcionamiento del concesionario se rige por las siguientes reglas:

- Sólo se venden los coches que en ese momento están en el concesionario. Esto implica que el día en que el número de coches en el concesionario es igual a cero, no puede realizarse ninguna venta.
- Cuando el concesionario compra un coche, lo pone a la venta el día siguiente.
- En el concesionario puede haber un máximo de 2 coches. Por consiguiente, no se realizará ninguna compra que haga que el número de coches sea superior a dos.

Este sistema puede modelarse como una Cadena de Markov, donde el estado del sistema es el número de coches que hay en el concesionario. Las transiciones entre estados son los cambios en el número de coches que se producen de un día para otro. El diagrama de transición del sistema se muestra en la Figura 7.1. La matriz de transición es la siguiente:

$$\mathbf{P} = \begin{pmatrix} 0.6100 & 0.3000 & 0.0900 \\ 0.1464 & 0.5572 & 0.2964 \\ 0.0400 & 0.2000 & 0.7600 \end{pmatrix} \quad (7.11)$$

Las probabilidades se han calculado considerando, para cada estado, los posibles eventos que pueden suceder a lo largo del día. El cálculo se ha realizado de la manera siguiente:

- $p_{0,0} = 0.61 \cdot 1$

Es la probabilidad de que, encontrándose el sistema en el estado "0", permanezca en dicho estado. Para ello, es preciso que no acuda ningún cliente que desee vender su coche. La probabilidad de que eso ocurra es igual a 0.61. Puesto que no hay vehículos en el concesionario, acuda el número de clientes que acuda con la intención de comprar, ese día el concesionario no podrá realizar ninguna venta.

- $p_{0,1} = 0.3 \cdot 1$
Es la probabilidad de que, encontrándose el sistema en el estado “0”, pase al estado “1”. Para ello es preciso que acuda un cliente con intención de vender su coche. La probabilidad de que ésto suceda es 0.3. Dado que los coches que compra el concesionario no son puestos a la venta hasta el día siguiente, no hay posibilidad de que concesionario venda ningún coche ese mismo día.
- $p_{0,2} = 0.09 \cdot 1$
Es la probabilidad de pasar de no tener ningún coche en stock, a tener dos. Para ello deben acudir 2 clientes que deseen vender su coche, lo cual se produce con una probabilidad igual a 0.09. Por la misma razón que en el caso anterior, ese día el concesionario no puede realizar ninguna venta.
- $p_{1,0} = 0.61 \cdot 0.24$
Es la probabilidad de pasar de tener un coche en stock a no tener ninguno. Para ello es preciso que no acuda ningún cliente con intención de vender su coche y que, a su vez, acuda un cliente que desee comprar un coche. La probabilidad del primer evento es 0.61, y la probabilidad del segundo es igual a 0.24. La probabilidad de que se den ambos simultáneamente, dado que son independientes entre sí, es igual al producto de la probabilidad de cada uno de ellos.
- $p_{1,1} = 0.61 \cdot 0.76 + 0.39 \cdot 0.24$
Es la probabilidad de que, habiendo un coche en stock, el balance de compra-venta del día haga que al cabo del mismo el concesionario siga teniendo un coche (el mismo o uno diferente). Para que esto suceda, es preciso que se de una de las dos siguientes posibilidades:
 - No acude ningún cliente, ni para comprar ni para vender. La probabilidad de que esto suceda es igual a $0.61 \cdot 0.76$.
 - Acuden 1 ó 2 clientes para vender su coche y además acuden 1 ó 2 clientes con intención de comprar un coche. Si acudieran 2 clientes a comprar, sólo el primero de ellos podría ser servido. Igualmente, si acudieran 2 clientes con intención de vender su coche, sólo se le compraría el coche al primero de ellos, ya que el concesionario no puede albergar más de 2 coches. La probabilidad de que se den simultáneamente estos eventos es igual a $0.39 \cdot 0.24$.
- Etc.

□

Ejemplo 7.3.2. La marca comercial AAA, que fabrica y comercializa un determinado producto, planea lanzar una campaña promocional con un doble objetivo.

- Por una parte, pretende reducir la probabilidad de que los clientes que en la actualidad consumen su producto dejen de hacerlo y pasen a consumir los productos de las dos marcas competidoras: B y C.
- Por otra parte, pretende aumentar la probabilidad de que los usuarios de las dos marcas de la competencia pasen a comprar los productos marca AAA.

Se supone que los clientes realizan la compra de producto cuando han consumido todo el que previamente habían comprado.

Cuando el cliente realiza la compra, puede suceder que

- siga consumiendo la marca que usaba anteriormente,
- o bien puede cambiar a una nueva marca.

Se han realizado unas estimaciones preliminares de cuál será la intención de compra del cliente, mientras esté realizándose la campaña publicitaria, en función de cual es la marca que actualmente consume. Las conclusiones obtenidas se resumen en la tabla siguiente:

| | Nueva marca: AAA | Nueva marca: B | Nueva marca: C |
|-------------------|------------------|----------------|----------------|
| Marca actual: AAA | 0.60 | 0.20 | 0.20 |
| Marca actual: B | 0.40 | 0.30 | 0.30 |
| Marca actual: C | 0.40 | 0.30 | 0.30 |

También se ha estimado el tiempo que tarda el cliente en consumir el producto de cada una de las tres marcas. Este tiempo es función de la marca ya que la cantidad de producto varía de una marca a otra. Se considera que éste es el tiempo que transcurre entre compras sucesivas:

| | Tiempo entre compras (días) |
|-----|-----------------------------|
| AAA | 35 ± 10 |
| B | 30 ± 8 |
| C | 25 ± 5 |

El objetivo del estudio es saber qué fracción del mercado tendrá la marca AAA 100 días después de iniciada la campaña promocional. Puede realizarse un modelo analítico (Cadena de Markov) o un modelo de simulación. Veamos ambos.

El problema puede modelarse como una Cadena de Markov si se realiza la simplificación de suponer que el cliente realiza 3 compras durante los 100 días. La matriz de transición de la Cadena de Markov es:

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix} \quad (7.12)$$

Inicialmente:

$$\pi^0 = \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right) \quad (7.13)$$

Después de 3 transiciones, se obtiene:

$$\pi^3 = \pi^0 \cdot \mathbf{P}^3 = (0.498 \quad 0.251 \quad 0.251) \quad (7.14)$$

El modelo de simulación puede incluir la variación en el tiempo entre compras. Realizando una simulación para 10000 clientes, después de 100 días, la marca AAA tiene el 50.2% del mercado, mientras que las marcas B y C tienen cada una el 24.9% del mercado. \square

7.4. Modelos analíticos de sistemas de colas

Los sistemas de colas (o “sistemas congestionados”, como se prefiera) son sistemas en los cuales existe la posibilidad de que una entidad demande un recurso que en ese instante no se encuentre disponible. Cuando se produce esa situación, la entidad debe esperar (en una cola) hasta que el recurso quede disponible.

Aunque los sistemas de colas normalmente se analizan mediante modelos de simulación, a menudo el conocimiento de los modelos analíticos es útil en el diseño y validación de los modelos de simulación.

En esta sección se consideran algunos modelos clásicos de teoría de colas, prestando especial atención a aquellas propiedades de estos modelos que pueden ser útiles para la comprensión de sistemas más complejos, que deben ser modelados usando modelos de simulación de eventos discretos.

Elementos de un sistema de colas

Todo sistema de colas consta de los tres elementos siguientes:

- Un proceso de llegada.
- Un proceso de servicio.
- Una disciplina de cola.

En un sistema de colas, el **proceso de llegada** se caracteriza por la distribución del tiempo que transcurre entre la llegada de una entidad y la llegada de la siguiente. Aunque se han desarrollado modelos analíticos para diferentes procesos de llegada, dos de los casos más importantes son cuando el intervalo entre llegadas está distribuido exponencialmente y cuando es constante.

La distribución exponencial juega un papel central en los modelos de colas, ya que es una distribución "sin memoria". Es decir, llamando t al intervalo entre llegadas sucesivas, y suponiendo que está distribuido exponencialmente, se satisface la siguiente igualdad:

$$\text{Prob}\{t > a + b | t > a\} = \text{Prob}\{t > b\} \quad (7.15)$$

Esta igualdad significa que la probabilidad de que se produzca una llegada durante las próximas b unidades de tiempo es independiente de cuando se produjo la última llegada. Por ejemplo, tanto si acaba de producirse una llegada como si se produjo hace 20 minutos, la probabilidad de que se produzca una llegada durante los próximos 5 minutos es la misma. Puede demostrarse que la única distribución de probabilidad sin memoria es la exponencial.

El tiempo medio entre llegadas puede ser constante o variar con el tiempo. En este último caso, el sistema debe ser analizado mediante un modelo de simulación.

El **proceso de servicio** se caracteriza por la distribución del tiempo necesario para servir (o procesar) a una entidad y por el número de servidores. Los modelos de teoría de colas a menudo asumen que el tiempo de servicio está distribuido exponencialmente, lo cual facilita el desarrollo de modelos analíticos. El número de servidores puede ser uno o varios. En algunos modelos el número de servidores es infinito.

Se define un parámetro, *intensidad del tráfico*, para medir la congestión del sistema:

$$\rho = \frac{\text{frecuencia de llegada de las entidades}}{\text{frecuencia a la que las entidades pueden ser procesadas}} \quad (7.16)$$

Para aquellos sistemas de colas en los cuales el tiempo medio de servicio y el tiempo medio entre llegadas son constantes, una definición equivalente del parámetro es:

$$\rho = \frac{\text{tiempo medio de servicio}}{(\text{tiempo medio entre llegadas}) \cdot (\text{número de servidores})} \quad (7.17)$$

Si la intensidad del tráfico es mayor que 1 (es decir, $\rho > 1$), la congestión del sistema crecerá sin límite con el tiempo.

La **disciplina de la cola** describe el orden en que las entidades captan el recurso. Disciplinas comunes son FIFO, colas de prioridad, el menor tiempo de servicio primero, selección aleatoria, etc. La disciplina de la cola también incluye características del sistema como la longitud máxima de la cola (cuando la cola adquiere ese número de elementos, se rechaza la llegada de nuevas entidades) y la impaciencia del cliente (los clientes esperando en la cola se impacientan y abandonan el sistema antes de ser servidos).

Los modelos de los sistemas de colas se designan de la forma siguiente:

$$A / B / s / K / E$$

donde:

A: Especifica el proceso de llegada.

B: Especifica el proceso de servicio.

s: Especifica el número de servidores.

K: Especifica el máximo número de entidades permitidas en el sistema.

E: Especifica la disciplina de la cola.

Los siguientes símbolos son comúnmente empleados en la clasificación de estos sistemas:

M: Tiempo de servicio o de llegada distribuido exponencialmente.

D: Tiempo de servicio o de llegada constante.

FIFO: Disciplina FIFO de la cola (First-In-First-Out).

SIRO: Disciplina aleatoria de la cola (Serve-In-Random-Order).

PRI: Disciplina basada en prioridades.

Ejemplo 7.4.1. Usando la terminología anterior, un sistema $M/D/3/50/PRI$ es un sistema de colas con un proceso de llegada exponencial, un tiempo de servicio constante, 3 servidores, un límite de 50 entidades en el sistema y una disciplina de cola en la cual las entidades captan los recursos de acuerdo a cierto criterio de prioridades. \square

Aunque existen múltiples medidas del comportamiento de un sistema de colas, las más comunes son las siguientes:

- Número medio de entidades en el sistema.
- Número medio de entidades en la cola.
- Tiempo medio que permanece una entidad en el sistema (=tiempo en cola + tiempo de servicio).
- Tiempo medio de espera en cola.
- Probabilidad de que haya exactamente i entidades en el sistema.
- Probabilidad de que el tiempo de espera sea superior a un determinado valor.

El modelo $M/M/1$

Los modelos de colas más sencillos son aquellos en los cuales los intervalos entre llegadas y los tiempos de proceso están distribuidos exponencialmente. Si se representa:

$\frac{1}{\lambda}$: el tiempo medio entre llegadas sucesivas (unidades: tiempo),

λ : la frecuencia de llegada de las entidades al sistema (unidades: 1/tiempo),

$\frac{1}{\mu}$: el tiempo medio de servicio (unidades: tiempo),

μ la frecuencia de servicio (unidades: 1/tiempo),

entonces:

$$\text{Prob}\{\text{Una llegada en el intervalo } [t, t + \Delta t]\} = \lambda \cdot \Delta t \quad (7.18)$$

$$\text{Prob}\{\text{Se completa el servicio en el intervalo } [t, t + \Delta t]\} = \mu \cdot \Delta t \quad (7.19)$$

Representando $P_n(t)$ la probabilidad de que haya exactamente n entidades en el sistema en el instante t , se obtiene:

$$\begin{aligned}
\underbrace{P_0(t + \Delta t)}_{\text{Probabilidad de que haya}} &= \underbrace{\mu_1(t) \cdot \Delta t \cdot P_1(t)}_{\text{Probabilidad de que haya}} \\
\underbrace{\text{cero entidades en el sistema}} &+ \underbrace{(1 - \lambda_0(t) \cdot \Delta t) \cdot P_0(t)}_{\text{Probabilidad de que haya}} \\
\underbrace{\text{en el instante } t + \Delta t}_{\text{en el instante } t + \Delta t} & \quad \text{una entidad en el instante } t \\
& \quad \text{y de que abandone el sistema} \\
& \quad \text{durante el intervalo } [t, t + \Delta t] \\
& \quad \text{cero entidades en el instante } t \\
& \quad \text{y de que no llegue ninguna durante} \\
& \quad \text{el intervalo } [t, t + \Delta t]
\end{aligned} \tag{7.20}$$

$$\begin{aligned}
\underbrace{P_n(t + \Delta t)}_{\text{Probabilidad de que haya}} &= \underbrace{\lambda_{n-1}(t) \cdot \Delta t \cdot P_{n-1}(t)}_{\text{Probabilidad de que haya}} \\
\underbrace{\text{n entidades en el sistema}} &+ \underbrace{\mu_{n+1}(t) \cdot \Delta t \cdot P_{n+1}(t)}_{\text{Probabilidad de que haya}} \\
\underbrace{\text{en el instante } t + \Delta t}_{\text{en el instante } t + \Delta t} & \quad \text{n-1 entidades en el instante } t \\
& \quad \text{y de que llegue una al sistema} \\
& \quad \text{durante el intervalo } [t, t + \Delta t] \\
& \quad \text{n+1 entidades en el instante } t \\
& \quad \text{y de que una abandone el sistema} \\
& \quad \text{durante el intervalo } [t, t + \Delta t] \\
& + \underbrace{(1 - \lambda_n(t) \cdot \Delta t) \cdot (1 - \mu_n(t) \cdot \Delta t) \cdot P_n(t)}_{\text{Probabilidad de que haya } n \text{ entidades en}} \\
& \quad \text{el instante } t \text{ y de que ni llegue ninguna ni} \\
& \quad \text{abandone ninguna el sistema durante} \\
& \quad \text{el intervalo } [t, t + \Delta t]
\end{aligned} \tag{7.21}$$

Se excluye la posibilidad de que se produzcan múltiples eventos en el intervalo $[t, t + \Delta t]$, por ejemplo, dos llegadas o tanto una llegada como un abandono, ya que se supone que la longitud del intervalo Δt es tan pequeña que la probabilidad de que se produzcan esos eventos es despreciable.

Reagrupando los términos y dividiendo por Δt se obtienen las expresiones siguientes:

$$\begin{aligned}
\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} &= \mu_1(t) \cdot P_1(t) \\
&- \lambda_0(t) \cdot P_0(t)
\end{aligned} \tag{7.22}$$

$$\begin{aligned}
\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} &= \lambda_{n-1}(t) \cdot P_{n-1}(t) \\
&+ \mu_{n+1}(t) \cdot P_{n+1}(t) \\
&- (\mu_n(t) + \lambda_n(t) - \lambda_n(t) \cdot \mu_n(t) \cdot \Delta t) \cdot P_n(t)
\end{aligned} \tag{7.23}$$

para $n : 1, 2, \dots$

Haciendo tender la longitud del intervalo a cero, se obtiene:

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= \mu_1(t) \cdot P_1(t) - \lambda_0(t) \cdot P_0(t) \\
\frac{dP_n(t)}{dt} &= \lambda_{n-1}(t) \cdot P_{n-1}(t) + \mu_{n+1}(t) \cdot P_{n+1}(t) \\
&- (\mu_n(t) + \lambda_n(t)) \cdot P_n(t) \quad \text{para } n : 1, 2, \dots
\end{aligned} \tag{7.24}$$

Si existe un límite en el número máximo de entidades que puede haber en el sistema, puede integrarse el sistema de ecuaciones diferenciales usando métodos numéricos. Las probabilidades obtenidas de esta solución son exactas (dentro de la precisión de la rutina numérica de integración), en comparación con las estimaciones obtenidas de los modelos de simulación.

Ejemplo 7.4.2. En una fábrica 2 técnicos de mantenimiento están a cargo de 8 máquinas. Las máquinas necesitan ser reparadas periódicamente: cuando una máquina está funcionando, el tiempo que transcurre entre averías está distribuido exponencialmente con media $\frac{1}{\lambda} = 1$ hora.

Si una máquina necesita ser reparada, pero los dos técnicos están ocupados, entonces la reparación de la máquina debe esperar a que uno de ellos quede libre. El tiempo que tarda el técnico en reparar la máquina está distribuido exponencialmente, con media $\frac{1}{\mu} = 0.5$ horas.

Las sucesivas peticiones de reparación que van surgiendo constituyen el proceso de llegada de entidades, que en este caso son las peticiones de servicio. El proceso de reparación constituye de proceso de servicio.

Para formular las ecuaciones diferenciales que describen el funcionamiento del sistema, se determinan en primer lugar las frecuencias de llegada y de servicio en función del número de máquinas, n , que se encuentran fuera de servicio.

- Si hay $n = 0$ máquinas fuera de servicio, la frecuencia de llegada de entidades (peticiones de servicio) es de $\lambda_0 = 8$ entidades por hora. La frecuencia de servicio (máquinas que pueden ser reparadas por hora) es $\mu_0 = 0$ entidades por hora, ya que todas ellas están funcionando.
- Si hay $n = 1$ máquina fuera de servicio, la frecuencia de llegada de entidades será $\lambda_1 = 7$ entidades por hora (ya que hay una ya estropeada), y la frecuencia de servicio será $\mu_1 = 2$ entidades por hora.
- Si hay n máquinas fuera de servicio, con $n \geq 2$, la frecuencia de llegada será $\lambda_n = 8 - n$ peticiones de servicio por hora, y la frecuencia de servicio será $\mu_n = 4$ máquinas reparadas por hora.

Los razonamientos anteriores pueden resumirse en la tabla siguiente:

| n | λ_n | μ_n |
|-----|-------------|---------|
| 0 | 8 | 0 |
| 1 | 7 | 2 |
| 2 | 6 | 4 |
| 3 | 5 | 4 |
| 4 | 4 | 4 |
| 5 | 3 | 4 |
| 6 | 2 | 4 |
| 7 | 1 | 4 |
| 8 | 0 | 4 |

Las ecuaciones diferenciales para este sistema se obtienen a partir de las ecuaciones (7.24), asignando valor a las frecuencias de llegada y servicio:

$$\frac{dP_0(t)}{dt} = 2 \cdot P_1(t) - 8 \cdot P_0(t) \quad (7.25)$$

$$\frac{dP_1(t)}{dt} = 8 \cdot P_0(t) + 4 \cdot P_2(t) - 9 \cdot P_1(t) \quad (7.26)$$

$$\begin{aligned} \frac{dP_n(t)}{dt} &= (8 - n + 1) \cdot P_{n-1}(t) + 4 \cdot P_{n+1}(t) \\ &- (4 + 8 - n) \cdot P_n(t) \quad \text{para } n : 2, \dots, 7. \end{aligned} \quad (7.27)$$

$$\frac{dP_8(t)}{dt} = 1 \cdot P_7(t) - 4 \cdot P_8(t) \quad (7.28)$$

donde P_n es la probabilidad de que haya n máquinas fuera de servicio. □

Modelos M/M/1 en el estacionario

Un sistema de colas se dice que está en estado *estacionario* cuando

$$\frac{dP_n(t)}{dt} = 0 \quad \text{para todo } n \quad (7.29)$$

Cuando el sistema está en estado estacionario, las ecuaciones (7.24) se pueden escribir de la forma siguiente:

$$\begin{aligned} \lambda_0 \cdot P_0 &= \mu_1 \cdot P_1 \\ (\lambda_n + \mu_n) \cdot P_n &= \lambda_{n-1} \cdot P_{n-1} + \mu_{n+1} \cdot P_{n+1} \quad n=1,2,\dots \end{aligned} \quad (7.30)$$

La intensidad del tráfico en el sistema, dado que hay un único servidor, es:

$$\rho = \frac{\lambda}{\mu} \quad (7.31)$$

La probabilidad de que haya exactamente n entidades en el sistema, P_n , se puede expresar en función de la intensidad del tráfico:

$$P_n = (1 - \rho) \cdot \rho^n \quad \text{para } n : 1, 2, \dots \quad (7.32)$$

El valor esperado del número de entidades en el sistema es:

$$L_s = \frac{\rho}{1 - \rho} \quad (7.33)$$

y el valor esperado del número de entidades en la cola es:

$$L_q = L_s - \rho = \frac{\rho^2}{1 - \rho} \quad (7.34)$$

Finalmente, el valor esperado del tiempo de espera de las entidades en la cola es:

$$W_q = \frac{\rho}{\mu \cdot (1 - \rho)} \quad (7.35)$$

En la Figura 7.2 se muestra la relación entre la longitud media de la cola y la intensidad de tráfico. Se observa que la longitud de la cola es relativamente pequeña hasta que la intensidad del tráfico llega a 0.60. Cuando la intensidad del tráfico crece por encima de ese valor, el tamaño de la cola crece rápidamente. Este comportamiento es típico de la mayoría de los sistemas de colas. A medida que la intensidad del tráfico se aproxima a 1.0, hay un rápido aumento de las medidas de la congestión, tales como L_q y W_q .

Ejemplo 7.4.3. *El cajero automático que ha instalado un banco en una de sus oficinas es usado por gran cantidad de clientes. El tiempo medio necesario para atender a un cliente es de 50 segundos, y el número de clientes que desean usar el cajero es 60 por hora. Tanto el tiempo de servicio como los intervalos entre llegadas sucesivas de los clientes son variables distribuidas exponencialmente.*

Varios clientes de han quejado de que han tenido que esperar en la cola un tiempo excesivo, por lo que solicitan que se instale un segundo cajero. Para determinar si se necesita una segunda máquina, el banco desea conocer la probabilidad de que un cliente deba esperar en la cola y el tiempo medio de espera de los clientes en la cola.

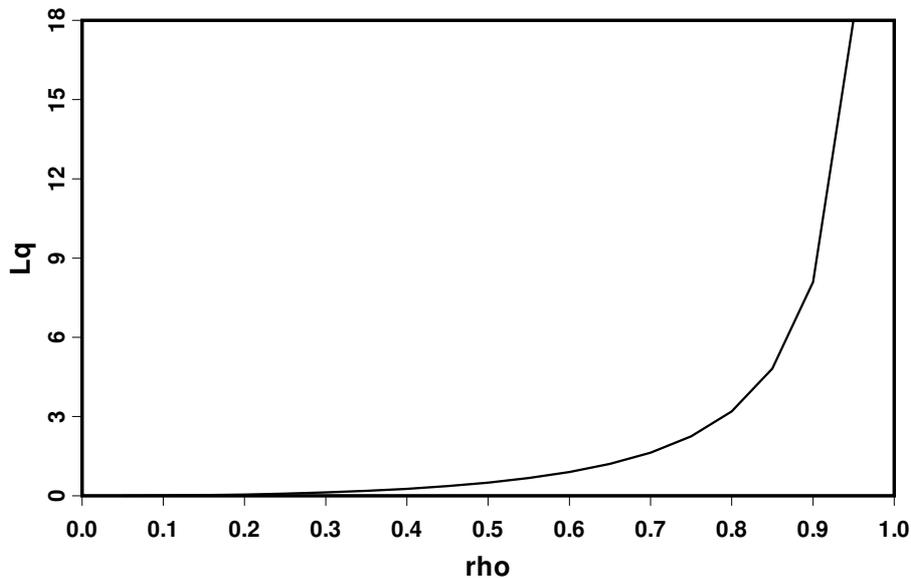


Figura 7.2: Número medio de entidades en cola vs intensidad del tráfico.

Para este sistema, la frecuencia de llegada es 60 clientes/hora, y la frecuencia de servicio es 72 clientes/hora. La intensidad del tráfico es: $\rho = 60/72 = 0.8333$. Usando las medidas del comportamiento del sistema M/M/1 en el estacionario:

$$P_0 = 1 - \rho = 0.1667 \quad (7.36)$$

$$W_q = \frac{\rho}{\mu \cdot (1 - \rho)} = \frac{0.8333}{72 \cdot 0.1667} = 0.0694 \text{ horas} = 4.167 \text{ minutos} \quad (7.37)$$

En consecuencia, aproximadamente el 83% de los clientes deben esperar en la cola, y el tiempo medio de espera son aproximadamente 4 minutos. Como medida adicional de la congestión del sistema, se puede calcular la longitud media de la cola:

$$L_q = \frac{\rho^2}{1 - \rho} = 4.2 \text{ clientes} \quad (7.38)$$

□

Parte III

Modelado y generación de las entradas aleatorias

Tema 8

Modelado de las distribuciones de probabilidad de entrada

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los conceptos fundamentales de la selección de las distribuciones de probabilidad de entrada.
- Aplicar las técnicas para la selección de las distribuciones usando datos experimentales y en ausencia de datos.
- Describir y comparar las distintas distribuciones de probabilidad.
- Analizar las entradas usando Arena.

8.1. Introducción

Es preciso disponer de datos con los que “alimentar” el modelo durante la simulación: tiempos entre llegadas consecutivas de las entidades, tiempos de proceso, tiempos entre fallos de los recursos, porcentaje de entidades que fallan cierto criterio, etc.

La adecuada selección de los datos de entrada es tan importante para el éxito del estudio como lo son la construcción del modelo o el posterior análisis de los resultados. En efecto, la decisión sobre la forma de generar las observaciones de cada variable aleatoria de entrada durante la simulación es una hipótesis de modelado, que se realiza basándose en consideraciones teóricas, datos medidos del sistema, etc. Aunque la estructura del modelo sea válida, si los datos de entrada no han sido recogidos y analizados adecuadamente, o no son representativos, entonces los resultados obtenidos de la simulación no serán correctos, y las decisiones tomadas en base a ellos serán erróneas.

La recogida de los datos experimentales del sistema es una de las tareas más costosas y delicadas en el proceso de modelado. El siguiente ejemplo pretende ilustrar esta afirmación.

Ejemplo 8.1.1. *Como parte de un curso en Modelado y Simulación, J. Banks y J. S. Carson propusieron como ejercicio a dos de sus alumnos la recogida de datos experimentales de un sistema real aparentemente sencillo¹: una lavandería, compuesta por 10 máquinas lavadoras y 6 secadoras, en la que los clientes lavan y secan su propia ropa.*

¹Este ejemplo está extraído de (Banks, Carson & Nelson 1996).

Sin embargo, la tarea de la recogida de datos pronto se convirtió en una tarea ingente. Los intervalos de tiempo entre llegadas de clientes no eran homogéneos: la distribución cambiaba en función de la hora del día, y del día de la semana. La lavandería abría 7 días a la semana, durante 16 horas al día, es decir, 112 horas a la semana. Con lo cual, era imposible cubrir completamente la operación de la lavandería con los limitados recursos de que se disponía (dos estudiantes, que además estaban matriculados de otras cuatro materias), y con las rígidas limitaciones de tiempo existentes (la simulación debía completarse en 4 semanas). Además, la distribución del tiempo entre llegadas durante una semana podía no reproducirse la siguiente semana. Como medida de compromiso, se seleccionó una muestra de tiempos, y las distribuciones del tiempo entre llegadas se calcularon y clasificaron de acuerdo a la frecuencia de llegadas como “alta”, “media” y “baja”.

La estimación de las distribuciones del tiempo de servicio también fue un problema difícil. Debía registrarse la proporción de clientes que demandaba cada tipo de servicio. El caso más simple eran los clientes que lavaban la ropa en una lavadora, y a continuación la secaban en una secadora. Sin embargo, había clientes que usaban dos lavadoras, seguidas de una única secadora, otros querían únicamente secar la ropa, etc. Dado que existía una dependencia entre la demanda de lavado y de secado para un determinado cliente, hubiera sido inapropiado tratar los tiempos de servicio para las lavadoras y secadoras separadamente, como variables aleatorias independientes.

Algunos clientes esperaban pacientemente a que finalizara el lavado o el secado de su ropa, y entonces sacaban la ropa de la máquina inmediatamente. Otros por el contrario, dejaban la máquina funcionando y volvían cuando calculaban que ésta había terminado. En las horas punta de afluencia de clientes, el encargado de la lavandería sacaba la ropa de estos clientes de la máquina y la dejaba metida en un cesto. En vista de ello, se decidió que la terminación del servicio se mediría en el instante en que la ropa se saca de la máquina, y no en el instante en que la máquina termina el proceso.

Asimismo, las máquinas se averiaban de cuando en cuando. El tiempo durante el cual permanecían fuera de servicio variaba desde un instante (cuando el encargado reparaba la máquina) a varios días (cuando era necesario sustituir alguna pieza). Los tiempos de las averías de corta duración eran registrados por los estudiantes, mientras que los tiempos de las averías de larga duración fueron estimados por el encargado. □

Cuando se dispone de muestras experimentales de la variable aleatoria de entrada, pueden aplicarse básicamente las tres técnicas siguientes:

- *Usar directamente los datos en la simulación.* Por ejemplo, si los datos representan el tiempo entre fallos de una máquina, entonces se usa uno de estos datos cada vez que se necesita un intervalo entre fallos en la simulación.
- *Definir una distribución empírica a partir de los datos experimentales.* Por ejemplo, si los datos representan el tiempo entre fallos, se muestrearán de esta distribución cuando se necesite un tiempo entre fallos en la simulación.
- *Ajustar una distribución teórica a los datos experimentales.* El proceso de ajuste consta típicamente de los tres pasos siguientes:
 1. En primer lugar se selecciona la familia de la distribución (exponencial, normal, de Poisson, etc.).
 2. A continuación se calcula qué valor de los parámetros de la familia de distribuciones proporciona el mejor ajuste a los datos. Con ello se selecciona una única distribución de la familia de distribuciones.
 3. Finalmente, se aplica un criterio para determinar la bondad del ajuste. Este criterio permite decidir, entre varias distribuciones, cuál es más representativa de los datos experimentales.

El procedimiento comúnmente más aceptado² consiste en tratar en primer lugar de modelizar la probabilidad de la variable de entrada mediante una distribución teórica, recurriendo a las distribuciones empíricas sólo como último recurso. Algunos de los motivos por los

²Ver, por ejemplo, (Pedgen et al. 1995) y (Law & Kelton 2000).

que puede ser preferible usar una distribución teórica, en lugar de una empírica, son los siguientes:

- La forma de las distribuciones empíricas puede ser considerablemente irregular, particularmente cuando se dispone de un número pequeño de datos. Sin embargo, en ocasiones la naturaleza del fenómeno que se está modelando no justifica estas irregularidades, que más bien son debidas a la propia naturaleza aleatoria de la recogida de datos del sistema. En estos casos puede ser preferible modelar el fenómeno “suavizando” los datos, mediante una distribución teórica.
- En aquellos casos en que existan fundamentos teóricos para modelar determinado fenómeno usando una determinada familia teórica de distribuciones, suele ser recomendable usar ese tipo de distribución teórica ajustada a los datos experimentales. No hay que olvidar que los datos experimentales recogidos son, en sí, aleatorios, con lo cual, una distribución empírica, obtenida a partir de un conjunto de observaciones, puede diferir mucho de la obtenida a partir de otro conjunto de observaciones del mismo proceso.
- Las distribuciones teóricas son una forma compacta de representar los datos. Por el contrario, si se dispone de n valores experimentales, la distribución empírica está definida mediante $2 \cdot n$ valores: los datos y las correspondientes probabilidades acumuladas. Si el conjunto de datos es grande, el almacenamiento de los datos puede ser una dificultad añadida.

No obstante, existen situaciones en que ninguna distribución teórica proporciona un ajuste adecuado a los datos. En estos casos suele ser preferible definir una distribución empírica en lugar de usar directamente los datos experimentales³. Esta última técnica presenta las dos desventajas siguientes:

- La simulación sólo puede reproducir lo que ha sucedido históricamente.
- Raramente se dispone del número de datos suficiente para realizar todas las réplicas (independientes) de la simulación que requiere el estudio.

En las secciones siguientes se discuten todos estos temas relativos a la selección de las distribuciones de entrada. La secuencia seguida en la exposición de las materias es la siguiente:

- En primer lugar, se realiza una introducción a las distribuciones de probabilidad: parametrización, distribuciones teóricas continuas y discretas más comúnmente empleadas, distribuciones empíricas, y desplazamiento y truncado de distribuciones.
- Seguidamente se discuten métodos estadísticos para contrastar la independencia y la homogeneidad de los datos experimentales.
- Las tres siguientes secciones están dedicadas a la descripción de las tareas propias del ajuste de distribuciones teóricas a los datos experimentales:
 - la selección de la familia de distribuciones,
 - la estimación de los parámetros, y
 - la medida de la bondad del ajuste realizado.
- A continuación, se discuten algunas consideraciones que pueden ser útiles para seleccionar las distribuciones de entrada cuando no se dispone de datos del sistema.
- Seguidamente, se describen algunos modelos probabilísticos de la llegada de las entidades al sistema.
- Finalmente, se muestra cómo emplear el analizador de entradas del entorno de modelado Arena para seleccionar las distribuciones de probabilidad de entrada del modelo.

³Durante determinadas fases de la validación del modelo, es aconsejable emplear directamente los datos del sistema real. Considérese el siguiente ejemplo: se desea validar el modelo de una oficina bancaria. El objetivo del estudio es estimar el tiempo medio de espera del cliente en la franja horaria de máxima actividad: entre las 8:30 y las 10:00 de la mañana. Se dispone de datos medidos en el sistema real de los intervalos entre llegadas de los clientes y de los tiempos de servicio, así como observaciones experimentales del tiempo medio de espera. Para validar el modelo puede ser útil emplear durante la simulación los datos medidos de los intervalos entre llegadas y de los tiempos de servicio (en lugar de generarlos de las distribuciones de probabilidad ajustadas), es decir, hacer que las entradas del modelo simulado sean iguales a los correspondientes valores medidos en el sistema real. De este modo, puede compararse más fielmente el comportamiento del modelo y del sistema real. La comparación puede realizarse empleando las técnicas que se describirán en temas posteriores, al tratar la comparación estadística de dos sistemas.

8.2. Distribuciones de probabilidad

Características más relevantes de las distribuciones teóricas

El Apéndice B, situado al final del libro, contiene información de interés práctico acerca de las distribuciones de probabilidad teóricas continuas más comúnmente usadas, y el Apéndice C acerca de las distribuciones discretas. Se recomienda en este punto leer ambos Apéndices, para saber qué tipo de información contienen y para qué tipo de consultas pueden resultar útiles.

Parámetros de las familias de distribuciones teóricas continuas

La densidad de probabilidad de cada una de las familias de distribuciones teóricas (por ejemplo, las distribuciones normales) posee unos determinados parámetros. La asignación de valor a estos parámetros permite identificar una distribución en concreto de entre la familia de distribuciones (por ejemplo, la distribución normal con media cero y varianza unidad). La mayoría de estos parámetros pueden clasificarse, de acuerdo con su interpretación física o geométrica, en uno de estos tres tipos: parámetros de *posición*, de *escala* o de *forma*.

- *Parámetro de posición*, γ . Especifica la posición en el eje de abscisas (eje x) del rango de valores de la distribución; normalmente, γ es el punto medio del rango o su extremo inferior. Una variación en γ supone un desplazamiento de la distribución, a la derecha o a la izquierda, sin variar ninguna otra de sus características.

Dos variables aleatorias, X e Y , de la misma familia de distribuciones, se dice que *difieren sólo en la posición* cuando existe un número real, γ , tal que $\gamma + X$ e Y tienen la misma distribución.

- *Parámetro de escala*, β . Determina la escala (o unidades) de medida de los valores en el rango de la distribución. Cuando γ está fijo a 0, un cambio en β comprime o expande la distribución sin alterar su forma básica.

Se dice que X e Y difieren sólo en la *escala* cuando existe algún real positivo, β , tal que $\beta \cdot X$ tiene la misma distribución que Y .

Se dice que X e Y difieren en la *posición* y en la *escala* cuando existen dos reales γ y β tales que $\gamma + \beta \cdot X$ tiene la misma distribución que Y .

- *Parámetro de forma*, α . Determina, con independencia de la posición y la escala, la forma básica de la distribución dentro de la familia general de distribuciones de interés. Un cambio en α generalmente altera las propiedades de la distribución más profundamente que un cambio en su posición o escala.

Algunas distribuciones no tienen ningún parámetro de forma (por ejemplo, la exponencial y la normal), mientras que otras poseen varios (por ejemplo, la distribución beta tiene dos). Si las distribuciones de X e Y poseen diferente valor de un parámetro de forma, no podrá hacerse que posean la misma distribución mediante un cambio de escala y de posición.

Distribuciones empíricas continuas

Los datos del sistema pueden ser usados para definir una *distribución empírica*. En el caso de las distribuciones empíricas continuas, el procedimiento de definición depende de si se dispone de las observaciones originales, o bien sólo se dispone de los datos agrupados (es decir, del número de observaciones que caen dentro de cada uno de una serie de intervalos en que se divide el rango de la variable aleatoria).

- Si se dispone de los datos originales, x_1, \dots, x_n , puede definirse, a partir de ellos, una función de probabilidad acumulada $F_X(x)$, continua y lineal a tramos. Para ello:

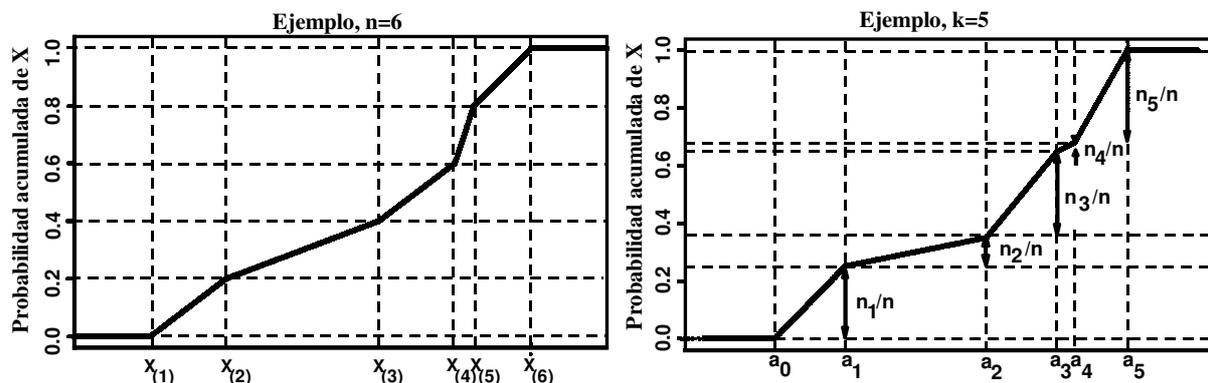


Figura 8.1: Distribuciones empíricas construidas a partir de datos individuales (izqda) y de los datos agrupados (drcha).

1. En primer lugar, deben ordenarse las observaciones crecientemente. La i -ésima menor muestra se representa de la forma siguiente: $x_{(i)}$. Es decir: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.
2. A continuación, se define la función de probabilidad acumulada (ver Figura 8.1) de la manera siguiente:

$$F_X(x) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \frac{i-1}{n-1} + \frac{x-x_{(i)}}{(n-1)(x_{(i+1)}-x_{(i)})} & \text{si } x_{(i)} \leq x < x_{(i+1)} \text{ para } i = 1, 2, \dots, n-1 \\ 1 & \text{si } x \geq x_{(n)} \end{cases} \quad (8.1)$$

Una desventaja de este tipo de distribución empírica es que las variables aleatorias generadas durante la simulación no podrán tener valores menores que $x_{(1)}$ ni mayores que $x_{(n)}$.

- Si se dispone de los datos agrupados en forma de histograma, puede aplicarse un método similar al anterior. Supóngase que los datos experimentales están agrupados en k intervalos adyacentes: $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$, de modo que el intervalo j -ésimo contiene n_j observaciones. El número total de observaciones es: $n_1 + n_2 + \dots + n_k = n$. Normalmente los valores a_j están equiespaciados, si bien el método para obtener la distribución empírica que se va a describir a continuación no requiere que se satisfaga esta condición.

El valor de la función de distribución acumulada, $F_X(x)$, puede definirse en los extremos de los intervalos de la forma siguiente (ver Figura 8.1):

$$F_X(a_0) = 0 \quad (8.2)$$

$$F_X(a_j) = \frac{n_1 + n_2 + \dots + n_j}{n} \text{ para } j : 1, 2, \dots, k \quad (8.3)$$

El valor de la función en los puntos internos a los intervalos puede calcularse mediante interpolación lineal de la manera siguiente:

$$F_X(x) = \begin{cases} 0 & \text{si } x < a_0 \\ F_X(a_{j-1}) + \frac{x-a_{j-1}}{a_j-a_{j-1}} (F_X(a_j) - F_X(a_{j-1})) & \text{si } a_{j-1} \leq x \leq a_j ; j : 1, 2, \dots, k \\ 1 & \text{si } a_k \leq x \end{cases} \quad (8.4)$$

Igual que en el caso anterior, las variables aleatorias generadas mediante estas distribuciones no pueden tomar valores menores que a_0 ni mayores que a_k .

Estos dos métodos son sólo dos de los muchos posibles para definir distribuciones empíricas de variables aleatorias continuas. Existe gran variedad de métodos, entre ellos cabe destacar aquellos que introducen en uno o a ambos lados una “cola exponencial” para eliminar el problema de la acotación en los valores de la variable aleatoria (ver, por ejemplo, (Bratley, Fox & Schrage 1987)).

Distribuciones empíricas discretas

Cuando se dispone de los datos, la distribución empírica discreta puede definirse de la manera siguiente: para cada posible valor x , la probabilidad $p(x)$ es la proporción de los datos que son iguales a x .

Cuando los datos son discretos, pero se dispone de ellos agrupados en forma de histograma, puede definirse la probabilidad de modo que la suma de las probabilidades $p(x)$ sobre todos los posibles valores de x en el intervalo sea igual a la proporción de las medidas que se encuentra en dicho intervalo. Si hay varios valores de x en el intervalo, el histograma no informa acerca de cómo repartir la probabilidad entre ellos, con lo cual debe hacerse arbitrariamente.

Desplazamiento y truncado de distribuciones

En algunas ocasiones debe redefinirse el *rango* (es decir, el conjunto de valores posibles) de las distribuciones teóricas, ya que de otro modo no representan adecuadamente el proceso. Considérese, por ejemplo, un servicio de atención al público en el que el tiempo requerido para atender al cliente se modela mediante una distribución gamma. El rango de la distribución gamma es $[0, \infty)$, sin embargo, en la práctica es imposible atender al cliente en menos de 30 segundos. La forma de recoger este hecho en el modelo consiste en modificar la distribución de probabilidad, de modo que la probabilidad de obtener tiempos de servicio inferiores a 30 segundos sea cero.

Existen al menos dos posibles formas de modificar el rango de las distribuciones:

- El primer método consiste en **desplazar** la distribución. Por ejemplo, para desplazar hacia la derecha la distribución $gamma(\alpha, \beta)$ una distancia $\gamma > 0$, debe sustituirse x por $(x - \gamma)$, es decir:

$$f_X(x) = \begin{cases} \frac{(x-\gamma)^{\alpha-1} \cdot e^{-\frac{x-\gamma}{\beta}}}{\beta^\alpha \cdot \Gamma(\alpha)} & \text{si } x > \gamma \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (8.5)$$

Si se emplea una distribución desplazada, el parámetro de posición γ deberá estimarse de la misma manera que los demás. La forma más adecuada de realizar esta estimación depende de la distribución. Por ejemplo, para la exponencial desplazada, es fácil estimar $\hat{\gamma}$ y $\hat{\beta}$ empleando el método de la máxima verosimilitud. Sin embargo, para la distribución gamma es preciso emplear métodos numéricos para estimar los 3 parámetros.

- El segundo método consiste en **truncar** la distribución. La densidad de probabilidad truncada al intervalo $[x_1, x_2]$, se representa como $f_X^*(x)$, y puede calcularse a partir de la densidad de probabilidad sin truncar, $f_X(x)$, de la forma siguiente:

$$f_X^*(x) = \begin{cases} \frac{f_X(x)}{a(x_1, x_2)} & \text{si } x \in [x_1, x_2], \text{ donde } a(x_1, x_2) = \int_{x_1}^{x_2} f_X(x) \cdot dx \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (8.6)$$

En general, la densidad de probabilidad truncada, $f_X^*(x)$, es discontinua en $x = x_1$ y $x = x_2$. Por ejemplo, en $x = x_1$ la densidad de probabilidad truncada salta de 0 a $f_X^*(x_1) = \frac{f_X(x_1)}{a(x_1, x_2)}$.

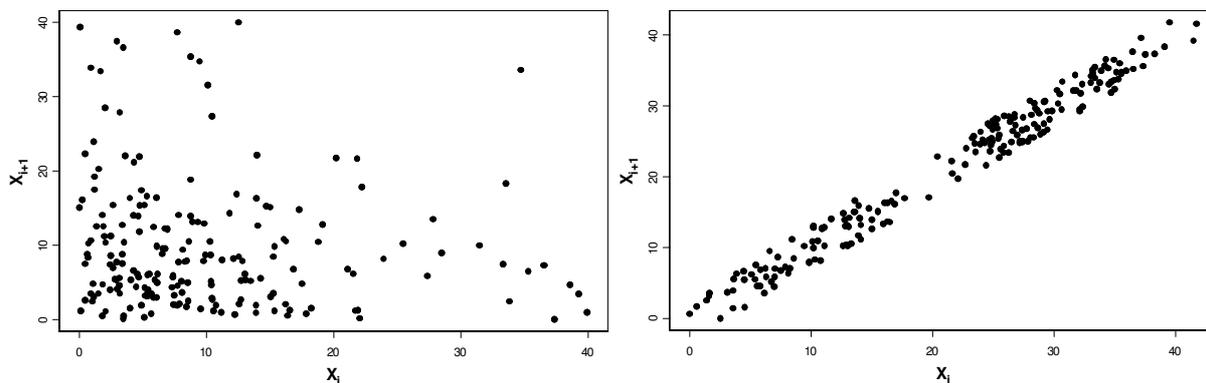


Figura 8.2: Ejemplo de evaluación de independencia de los datos.

Un método práctico de decidir a partir de qué valor o valores debe truncarse la distribución es mediante la inspección visual del histograma de las observaciones experimentales.

8.3. Análisis de la independencia y homogeneidad de los datos experimentales

Algunas de las técnicas estadísticas que se aplican en el ajuste de distribuciones a los datos experimentales asumen que éstos son independientes entre sí (por ejemplo, la estimación de máxima verosimilitud y el test chi-cuadrado). Por consiguiente, si las muestras no son *independientes*, no es correcto aplicar estas técnicas (aunque siempre podrán aplicarse técnicas heurísticas como son los histogramas, etc.). Desde este punto de vista, resulta de gran importancia saber contrastar la independencia de las observaciones experimentales.

Por otra parte, en ocasiones se recogen diferentes conjuntos independientes de observaciones de determinado fenómeno aleatorio y es preciso saber si estos conjuntos de datos son *homogéneos* (puede considerarse idénticamente distribuidos entre sí) y por tanto se pueden juntar en una única muestra. Para ello es preciso conocer técnicas para contrastar la *homogeneidad* de varios conjuntos de observaciones.

Análisis de la independencia de los datos experimentales

Los datos experimentales presentan en ocasiones una correlación temporal. Por ejemplo, si se mide experimentalmente el tiempo que deben esperar los clientes en una cola, se comprobará que las observaciones están correlacionadas. Por ejemplo, si un cliente ha debido esperar mucho tiempo para ser atendido (el sistema está congestionado), lo normal es que el cliente situado tras él en la cola también haya debido esperar mucho tiempo.

Considérese una secuencia de datos experimentales x_1, x_2, \dots, x_n , dispuestos en el mismo orden que han sido recogidos. Una técnica gráfica para analizar si los datos de esta secuencia son independientes consiste en representar las parejas de puntos (x_i, x_{i+1}) , para $i = 1, 2, \dots, n$. Si los datos son independientes, los puntos estarán distribuidos aleatoriamente, mientras que si están correlacionados, tenderán a alinearse.

Ejemplo 8.3.1. Como ilustración de aplicación de esta técnica, en la parte izquierda de la Figura 8.2 se han representado 200 observaciones independientes e idénticamente distribuidas

(aproximadamente de forma exponencial con media 10). En la parte derecha de la Figura 8.2 se muestran 200 observaciones correlacionadas entre sí. \square

Esta técnica gráfica es sólo una de las múltiples técnicas que existen. Se volverá sobre este tema más adelante, al discutir los tests para contrastar la independencia de las secuencias de números pseudo aleatorios.

Análisis de la homogeneidad de los datos experimentales

Supóngase que se han recogido k conjuntos independientes de muestras, donde posiblemente cada conjunto tiene un número diferente de muestras. Por ejemplo, el conjunto de muestras i -ésimo está compuesto por las n_i observaciones siguientes: $x_{i,1}, x_{i,2}, \dots, x_{i,n_i}$. A su vez, las muestras que componen cada uno de los conjuntos son independientes entre sí.

Se pretende contrastar la hipótesis siguiente:

H_0 : Las distribuciones de todos los conjuntos de muestras son idénticas.

H_1 : Al menos la distribución de uno de los conjuntos de muestras tiende a producir valores mayores que al menos la distribución de otro de los conjuntos de muestras.

Para contrastar esta hipótesis nula puede emplearse el *test de Kruskal-Wallis*. Se trata de un test *no paramétrico*, ya que no realiza ninguna hipótesis sobre la distribución de los datos. Consta de los siguientes tres pasos:

1. Ordenar el conjunto completo de muestras de menor a mayor y asignar a cada muestra su número de orden (entre 1 y $n = \sum_{i=1}^k n_i$). Sea $R(x_{i,j})$ el número de orden de la muestra j del conjunto i .
2. Para cada conjunto de muestras, calcular la suma de los números de orden que se ha asignado a sus muestras:

$$R_i = \sum_{j:1}^{n_i} R(x_{i,j}) \text{ para } i : 1, \dots, k \quad (8.7)$$

3. El estadístico del test se define de la manera siguiente⁴:

$$T = \frac{12}{n \cdot (n + 1)} \cdot \sum_{i:1}^k \frac{R_i^2}{n_i} - 3 \cdot (n + 1) \quad (8.8)$$

Se rechaza la hipótesis nula con nivel α si se satisface $T > \chi_{k-1, 1-\alpha}^2$, donde $\chi_{k-1, 1-\alpha}^2$ es el punto crítico $1 - \alpha$ de la distribución chi-cuadrado con $k - 1$ grados de libertad.

8.4. TAREA I: Selección de la familia de distribuciones

Supongamos que se pretende emplear una distribución de probabilidad teórica para modelar una entrada aleatoria de la simulación. El primer paso en la selección de la distribución teórica es decidir qué familia de distribuciones parece más apropiada. Esta decisión se toma en base al análisis de la forma de la distribución, sin tener en cuenta, por el momento, el valor de los parámetros de la distribución.

En esta sección se describen algunas técnicas para investigar qué familia de distribuciones permite representar más adecuadamente una determinada variable aleatoria de entrada.

⁴Este estadístico es válido sólo cuando no hay dos muestras de igual valor. En caso contrario, debe usarse un estadístico diferente (ver la referencia citada en (Law & Kelton 2000)).

Consideraciones teóricas

En ocasiones, se dispone a priori del suficiente conocimiento acerca del proceso a modelar como para poder seleccionar una familia de distribuciones, o al menos para poder descartar algunas otras. A continuación se discuten varias recomendaciones sobre el uso de algunas distribuciones teóricas continuas.

- *Uniforme*. Resulta adecuada cuando la entrada toma con igual probabilidad cualquier valor de cierto intervalo. Se usa como primera aproximación⁵ para una variable que se supone que varía aleatoriamente en determinado intervalo, y de la que no se sabe nada más.
- *Triangular*. Se usa como primera aproximación en ausencia de datos, cuando se pretende representar un proceso para el cual pueden establecerse unos límites (*mínimo* y *máximo*) y el valor más probable (*modo*).
- *Gamma*. Suele emplearse para modelar el tiempo necesario para completar una tarea. Esta distribución puede tomar una gran variedad de formas, en función del valor de su parámetro de forma.
- *Exponencial*. Se usa para representar el intervalo de tiempo entre eventos independientes, por ejemplo, la llegada de entidades al sistema. Si el número de eventos por intervalo de tiempo es una variable aleatoria discreta de Poisson de media λ , entonces los intervalos de tiempo entre llegadas están distribuidos exponencialmente con media $\beta = \frac{1}{\lambda}$. Cuando la variable aleatoria representa un intervalo de tiempo, la distribución exponencial posee una propiedad única: no tiene memoria.
- *Weibull*. Se usa en los modelos de fiabilidad, para representar los tiempos de vida (tiempo hasta el fallo) de dispositivos en los cuales el uso afecta a su tiempo esperado de vida (dispositivos eléctricos, muelles, etc.). En ocasiones, se usa también para representar el tiempo necesario para completar una tarea.
- *Normal*. Se emplea para describir fenómenos que son suma o promedio de una gran cantidad de variables aleatorias independientes entre sí, y tales que cada una de ellas realiza una pequeña contribución a la suma (*Teorema del Límite Central*). Por ejemplo, si se reemplazan los tiempos de proceso de varios subprocesos independientes por un único tiempo de proceso, que es la suma de aquellos, el tiempo de proceso resultante tiene una distribución que se aproxima a la normal al aumentar el número de subprocesos⁶. Esto se cumple con independencia de qué distribuciones de probabilidad tengan los tiempos de los subprocesos.

Seguidamente, se proporcionan recomendaciones sobre el uso de algunas distribuciones teóricas discretas:

- *Bernoulli*. La variable aleatoria representa el resultado de un experimento con dos únicos posibles resultados. Este tipo de experimentos se conocen como “*experimentos de Bernoulli*”.
- *Uniforme discreta*. La variable aleatoria es el resultado de un experimento con varios posibles resultados, todos ellos igualmente probables.
- *Binomial*. Es adecuada cuando se pretende estudiar el número de miembros de un grupo de individuos que poseen cierta característica. El grupo de individuos ha sido escogido aleatoriamente, y posee un tamaño fijo, conocido a priori. Cada individuo tiene

⁵En algunos textos, por ejemplo (Pedgen et al. 1995) se califica la distribución uniforme como la “distribución de máxima ignorancia”, ya que se usa cuando no se tiene absolutamente ninguna idea de la forma de la distribución. Su empleo se justifica mediante la afirmación siguiente: “Dado que no tengo ninguna razón para pensar lo contrario, asumiré igual probabilidad para todos los posibles resultados”.

No obstante, el empleo de la distribución uniforme no está exento de polémica. Hay autores como Pedgen et al. (1995) que afirman no haber encontrado ningún proceso real (y ni tan siquiera ser capaces de imaginarlo) que esté distribuido uniformemente. En su lugar, recomiendan el uso de la distribución triangular, con la media en el punto central del rango, o incluso una distribución normal cuya media sea el punto central del rango y la desviación estándar un cuarto del rango (es decir, $\text{rango} = 4 \cdot \sigma$). Si se usa la distribución normal, es preciso truncarla para evitar obtener valores demasiado grandes o demasiado pequeños.

⁶A menos que los tiempos de proceso individuales tengan distribuciones muy sesgadas (es decir, muy poco simétricas), esta aproximación es frecuentemente razonable para un número de subprocesos tan pequeño como cuatro (Pedgen et al. 1995).

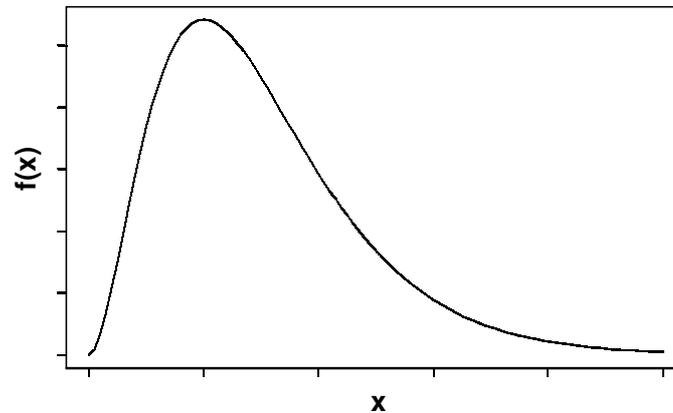


Figura 8.3: Forma típica de la densidad de probabilidad.

la misma probabilidad de poseer la característica. El hecho de que un individuo posea la característica es independiente del hecho de que otro la posea: saber que ciertos individuos poseen la característica no ayuda a decidir si otro individuo en concreto la tiene.

- *Geométrica*. Es aplicable cuando se pretende estudiar el número de repeticiones de un experimento de Bernoulli que deben hacerse hasta que se produce uno determinado de los dos posibles resultados.
- *Negativa binomial*. Es adecuada para describir el número de experimentos de Bernoulli “fallidos” antes de que se dé, un determinado número de veces, uno determinado de los dos posibles resultados.
- *Poisson*. Es adecuada para modelizar el número de eventos aleatorios que suceden en un intervalo de tiempo fijo.

Estadísticos

Los valores de algunos estadísticos de la muestra experimental, como son la media, la mediana, la varianza, etc. pueden guiar en la elección de la familia de la distribución. En cualquier caso, debe tenerse presente que el valor del estadístico calculado de la muestra es una observación aleatoria, es decir, el valor calculado no coincide (en general) con el valor verdadero del estadístico de la distribución. En conclusión, estos métodos tienen sólo carácter orientativo y en ningún caso proporcionan información definitiva. A continuación se muestran algunos ejemplos de aplicación⁷:

- Si la media y la mediana de los datos experimentales difieren significativamente entre sí, es indicación de que probablemente la distribución de entrada no sea simétrica.
- El *coeficiente de variación*, definido como el cociente de la desviación estándar y la media, $cv = \frac{\sqrt{\sigma^2}}{\mu}$, puede proporcionar información muy útil acerca del tipo de distribución.

Dado un conjunto de observaciones, x_1, \dots, x_n , el coeficiente de variación puede estimarse operando los estimadores de la media y la varianza de la forma siguiente: $\frac{\sqrt{S^2(n)}}{\bar{X}(n)}$. Estos

últimos se calculan de la forma siguiente: $\bar{X}(n) = \frac{1}{n} \cdot \sum_{i:1}^n x_i$ y $S^2(n) = \frac{\sum_{i:1}^n (x_i - \bar{X}(n))^2}{n-1}$.

Las siguientes propiedades hacen del coeficiente de variación una herramienta útil:

⁷Puede encontrarse una discusión detallada en (Law & Kelton 2000).

- El coeficiente de variación de una distribución exponencial vale $cv = 1$, con independencia del valor de la media. Si el valor del coeficiente obtenido de la muestra experimental está próximo a uno, tiene fundamento realizar la hipótesis de que la distribución es exponencial.
- Supóngase que la representación gráfica de la densidad de probabilidad de los datos experimentales tiene una forma similar a la mostrada en la Figura 8.3. En esta situación, cabe preguntarse qué tipo de distribución será más adecuada. Por ejemplo, la distribución gamma o la distribución de Weibull tienen esa forma cuando el parámetro de forma es mayor que uno. También tiene una forma similar a la de la Figura 8.3 la distribución lognormal. En este caso, el coeficiente de variación puede servir de orientación, ya que:
 - Las distribuciones gamma y Weibull con la forma de la Figura 8.3, es decir, con el parámetro de forma mayor que uno, satisfacen: $cv < 1$.
 - La distribución lognormal siempre tiene una apariencia similar a la de la Figura 8.3, y su coeficiente de variación puede ser cualquier número real positivo.
 Por consiguiente, si la distribución tiene una forma similar a la de la Figura 8.3 y el valor del coeficiente de variación estimado de los datos experimentales es mayor que uno, entonces es recomendable usar la distribución lognormal frente a la distribución gamma o la distribución de Weibull.
- El estadístico $\tau = \frac{\sigma^2}{\mu}$ juega para las distribuciones discretas un papel similar al coeficiente de variación para las distribuciones continuas. Igualmente, puede calcularse una estimación a partir de los datos experimentales de la forma siguiente: $\hat{\tau} = \frac{S^2(n)}{X(n)}$. Este estadístico es útil para discriminar entre:
 - Poisson. Tiene $\tau = 1$.
 - Binomial. Tiene $\tau < 1$.
 - Binomial negativa. Tiene $\tau > 1$.
- El *coeficiente de simetría* o *sesgo*, ν , proporciona una medida de la simetría de la distribución. En las distribuciones simétricas, como la normal, el sesgo vale cero. Se define como el tercer momento de la distribución dividido por el cubo de la desviación estándar:

$$\nu = \frac{E\{(X - \mu)^3\}}{\sigma^3} \quad (8.9)$$

Dado un conjunto de observaciones, x_1, \dots, x_n , puede estimarse el valor del sesgo de la forma siguiente:

$$\hat{\nu} = \frac{\hat{m}_3(n)}{(S^2(n))^{3/2}} \quad \text{donde } \hat{m}_3(n) = \frac{\sum_{i=1}^n (x_i - \bar{X}(n))^3}{n} \quad (8.10)$$

Histogramas

El histograma es básicamente una estimación gráfica de la densidad de probabilidad de los datos experimentales, x_1, \dots, x_n . Las funciones densidad de probabilidad de las distribuciones teóricas tienen formas reconocibles, con lo cual, el histograma sirve de ayuda para seleccionar qué familia o familias de distribuciones tiene sentido emplear para modelar el proceso de entrada. Para ello debe compararse la forma del histograma con la forma de la distribución teórica, sin considerar la escala o la posición.

Para construir el histograma de una variable aleatoria continua deben seguirse los siguientes dos pasos:

1. Dividir el rango de valores de los datos experimentales en k intervalos disjuntos, adyacentes y de igual longitud: $[b_0, b_1), [b_1, b_2), \dots, [b_{k-1}, b_k)$. Cada uno de estos intervalos se denomina una *clase* o *categoría*. A continuación se agrupan los datos pertenecientes a cada clase.

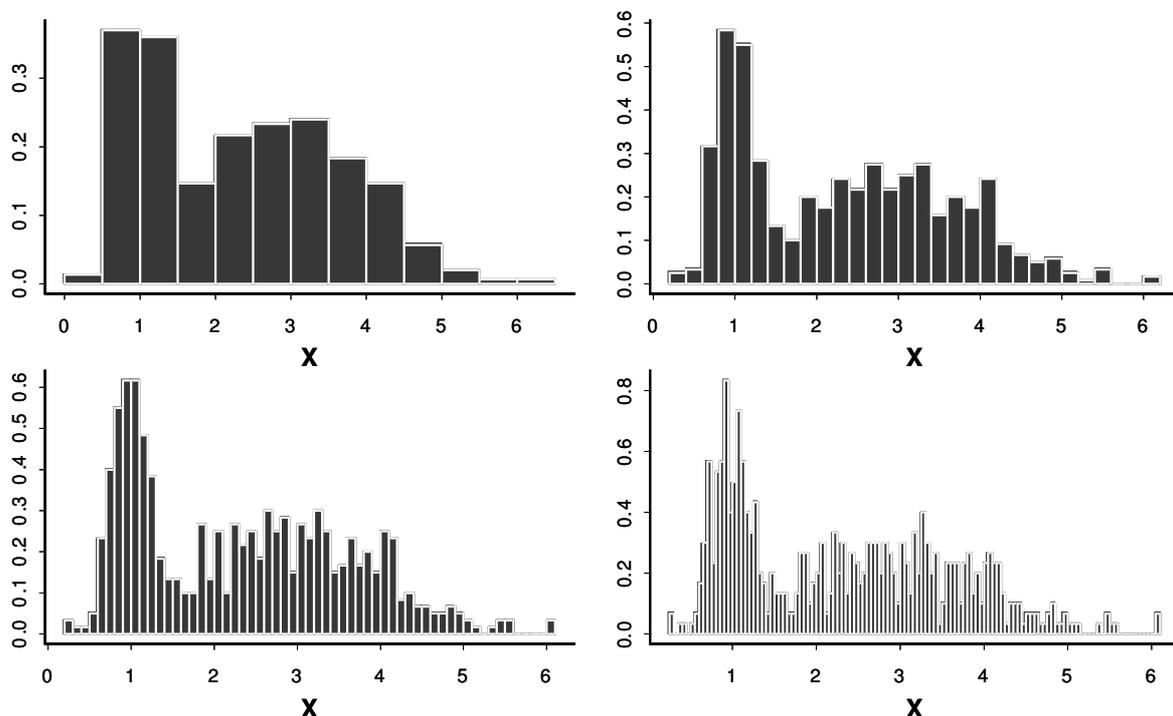


Figura 8.4: Histogramas de un mismo conjunto de datos.

2. Sea q_j la frecuencia de la clase j , es decir, la proporción de los datos experimentales, x_1, \dots, x_n , que se encuentran en el intervalo $[b_{j-1}, b_j)$. El histograma es la siguiente función constante a tramos:

$$h(x) = \begin{cases} 0 & \text{si } x < b_0 \\ q_j & \text{si } b_{j-1} \leq x < b_j, \text{ para } j = 1, \dots, k \\ 0 & \text{si } b_k \leq x \end{cases} \quad (8.11)$$

Dado que el acto de agrupar los datos destruye muchos de sus detalles originales, la elección del número de clases del histograma es importante. Existen varias recomendaciones⁸ y reglas. Una de ellas es la *regla de Sturges*, que dice que el número de clases, k , debe escogerse de acuerdo a la fórmula siguiente⁹:

$$k = \left\lceil 1 + \log_2(n) \right\rceil = \left\lceil 1 + 3.322 \cdot \log_{10}(n) \right\rceil \quad (8.12)$$

Otra recomendación alternativa (Law & Kelton 2000) consiste probar diferentes anchuras de intervalo y escoger aquella que proporciona un histograma “suave”, con una forma similar a alguna distribución teórica. Como puede observarse, la construcción de los histogramas tiene cierta componente subjetiva, lo cual constituye su principal desventaja. En el caso

⁸En (Pedgen et al. 1995) se ofrecen las recomendaciones siguientes. El número de clases normalmente varía entre 5 y 20, en función de los datos; lo más habitual es usar entre 8 y 12 clases. Es recomendable tener al menos 3 ó 5 observaciones en cada clase.

⁹Se emplean los siguientes símbolos para representar la aproximación de números reales a valores enteros:
 $\lfloor x \rfloor$ representa el número entero más grande tal que es $\leq x$.
 $\lceil x \rceil$ representa el número entero más pequeño tal que es $\geq x$.

discreto, la construcción del histograma (o *diagrama de barras*) es más sencilla ya que no es preciso definir ni los intervalos ni agrupar los datos.

En la Figura 8.4 se muestran cuatro histogramas, con diferente número de clases, de los mismos 600 datos experimentales. Estos datos corresponden a medidas realizadas del tiempo necesario para atender a un cliente. Los histogramas se han reescalado de modo que representen una densidad de probabilidad: la suma del área de las barras es igual a uno. Se observa que el histograma tiene dos *modos* (máximos de la densidad de probabilidad) locales, uno entorno a 1 minuto y otro entorno a 3 minutos. Esto implica que existen esencialmente dos tipos de servicio, con distinta duración, que puede requerir el cliente.

Al ajustar los datos experimentales debe tenerse en cuenta esta “bimodalidad”. Las observaciones pueden separarse en dos grupos, siendo p_j la proporción de observaciones del grupo j (con $j = 1, 2$). Las observaciones del grupo j se usan para ajustar la distribución teórica $f_j(x)$. La densidad de probabilidad global, que considera ambos tipos de servicio, se calcula, a partir de las dos densidades anteriores, de la forma siguiente:

$$f(x) = p_1 \cdot f_1(x) + p_2 \cdot f_2(x) \quad (8.13)$$

Gráficas cuantil-cuantil

La *probabilidad acumulada* (también llamada *función de distribución acumulada*) de una variable aleatoria X se define de la forma siguiente: $F_X(x) = \text{Prob}\{X \leq x\}$. Por ello, si los datos experimentales x_1, \dots, x_n están distribuidos de la misma forma que la variable aleatoria X , una aproximación razonable a $F_X(x)$ es la proporción de los datos experimentales que son menores o iguales que x .

Cabría pues, plantearse comparar la distribución de probabilidad acumulada obtenida a partir de los datos experimentales, con las distribuciones de probabilidad acumulada de las distribuciones estándar. Sin embargo, éstas suelen tener forma de “S”, con lo cual la comparación visual no suele ser demasiado esclarecedora.

Existen técnicas para reducir el problema de la comparación de funciones distribución de probabilidad acumulada a decidir cuál, de entre varias gráficas, se asemeja más a una recta. Una de estas técnicas, denominada *gráfica cuantil-cuantil* o *gráfica Q-Q*, está basada en la comparación de los cuantiles o puntos críticos de las distribuciones continuas.

El cuantil q (con $0 < q < 1$) de una distribución es un número x_q que satisface:

$$F_X(x_q) = q \quad (8.14)$$

Representando F_X^{-1} a la función inversa de la probabilidad acumulada, una definición equivalente del cuantil es:

$$x_q = F_X^{-1}(q) \quad (8.15)$$

La técnica gráfica Q-Q se basa en las dos propiedades siguientes:

- Si las variables aleatorias X e Y están igualmente distribuidas, entonces sus densidades de probabilidad y sus cuantiles son iguales: $y_q = x_q$. La representación gráfica de los puntos (x_q, y_q) , para $q \in (0, 1)$, es una recta con pendiente unidad que pasa por el origen.
- Si las variables aleatorias X e Y pertenecen a la misma familia de distribuciones y sus densidades acumuladas difieren en el valor los parámetros de posición (γ) y escala (β), es decir, $F_Y(y) = F_X\left(\frac{y-\gamma}{\beta}\right)$, entonces la relación entre los cuantiles de las distribuciones es: $y_q = \gamma + \beta \cdot x_q$. Consiguientemente, la representación gráfica de los puntos (x_q, y_q) , para $q \in (0, 1)$, es una recta con pendiente β que no pasa por el origen.

A la vista de las dos propiedades anteriores, la representación de los pares de puntos (x_q, y_q) , es una herramienta útil para comparar dos distribuciones. En particular, para com-

parar la distribución empírica obtenida de los datos experimentales con una distribución teórica. A efectos de la aplicación de la técnica, sólo es relevante el parámetro de forma de la distribución teórica. Los parámetros de escala y posición son irrelevantes, con lo cual pueden escogerse de la forma que resulte más sencilla.

Los dos pasos a seguir para comparar la distribución empírica de un conjunto de datos experimentales, x_1, \dots, x_n , con la distribución F_X son los siguientes:

1. En primer lugar hay que construir una función de probabilidad acumulada empírica, \tilde{F} , a partir de los datos experimentales. Para ello es preciso ordenar crecientemente los datos experimentales: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, donde $x_{(i)}$ representa el dato que ocupa la i -ésima posición. El valor de la distribución empírica en cada uno de los datos experimentales es igual al número de datos menor o igual que éste:

$$\tilde{F}(x_{(i)}) = \frac{i}{n} \text{ para } i : 1, \dots, n \quad (8.16)$$

Esta definición presenta la desventaja de que la probabilidad acumulada vale 1 para el valor experimental mayor: $\tilde{F}(x_{(n)}) = \frac{n}{n} = 1$. Una forma de evitar este inconveniente es modificar ligeramente la definición (Law & Kelton 2000):

$$\tilde{F}(x_{(i)}) = \frac{i - 0.5}{n} \text{ para } i : 1, \dots, n \quad (8.17)$$

2. A continuación hay que construir la gráfica Q-Q. Puesto que $x_q = x_{(i)}$ es el cuantil $q = \frac{i-0.5}{n}$ de la distribución empírica, la gráfica consiste en la representación de los puntos:

$$\left(x_{(i)}, F_X^{-1} \left(\frac{i - 0.5}{n} \right) \right) \text{ para } i : 1, \dots, n \quad (8.18)$$

Si estos puntos se ajustan razonablemente bien a una recta (con independencia de su pendiente o de si pasa o no por el origen), entonces está justificada la hipótesis de que la familia y el factor de forma de la distribución de la que están muestreados los datos experimentales y de F_X coinciden.

En la Figura 8.5 se muestra un ejemplo sencillo de aplicación de esta técnica. Se ha recogido una muestra de 100 datos experimentales. En primer lugar se dibuja el histograma de los datos. Se observa que el histograma es aproximadamente simétrico (esta observación puede contrastarse estimando el sesgo a partir de los datos experimentales). Por ello, tiene sentido realizar comparaciones con la distribución normal o con la distribución beta con $\alpha_1 = \alpha_2$. Como puede observarse de la segunda gráfica Q-Q, la comparación con la distribución normal es bastante satisfactoria. Sobre las gráficas Q-Q se ha dibujado la recta ajustada en cada caso. Como referencia del comportamiento de la gráfica Q-Q cuando la distribución empírica y teórica tienen una forma considerablemente diferente, se han comparado los datos experimentales con la distribución uniforme y exponencial (primera y tercera gráfica Q-Q respectivamente).

8.5. TAREA II: Estimación de los parámetros

Una vez seleccionada una familia de distribuciones en la Tarea I, hay que escoger aquella distribución en concreto de la familia que mejor se ajusta a los datos experimentales. La forma de seleccionar la distribución es asignar valor a los parámetros de la distribución. Existen diferentes métodos de estimar el valor de los parámetros. En esta sección discutiremos sólo uno de ellos: los *estimadores de máxima verosimilitud*¹⁰.

¹⁰En (Law & Kelton 2000) se discuten las ventajas de esta técnica frente a otras. También puede encontrarse la explicación de propiedades importantes del estimador de máxima verosimilitud.

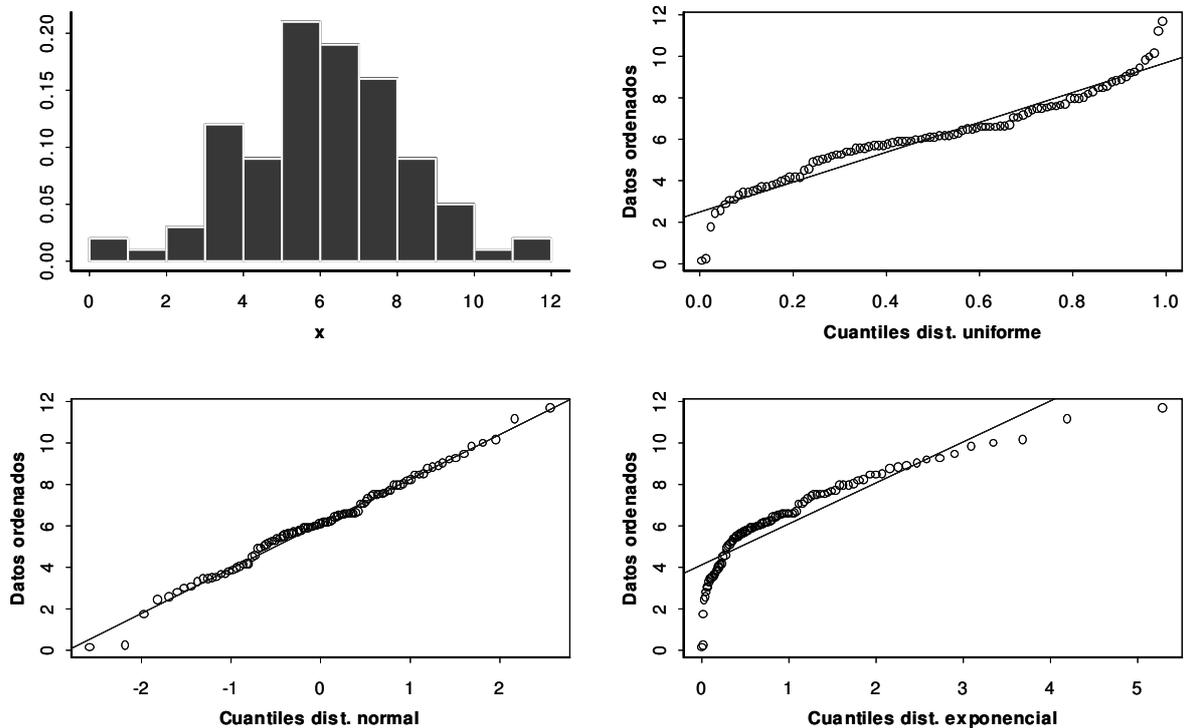


Figura 8.5: Ejemplo de aplicación de las gráficas Q-Q.

Estimadores de máxima verosimilitud

Los fundamentos del método se comprenden mejor en el **caso discreto**. Se pretende ajustar a los datos experimentales la probabilidad $p_X(x, \theta)$ de cierta distribución discreta, donde θ es el parámetro cuyo valor quedará determinado por el ajuste.

La *función verosimilitud*, $L(\theta)$, se define a partir de los datos experimentales x_1, \dots, x_n (los mismos usados en la Tarea I) de la manera siguiente:

$$L(\theta) = p_X(x_1, \theta) \cdot p_X(x_2, \theta) \cdot \dots \cdot p_X(x_n, \theta) \quad (8.19)$$

Puesto que los datos x_1, \dots, x_n son independientes, $L(\theta)$ es igual a la probabilidad de haber obtenido ese conjunto de datos al realizar n observaciones independientes de la distribución. Tal como indica la representación $L(\theta)$, el valor de la función de verosimilitud L depende del valor del parámetro θ .

Se define el *estimador de máxima verosimilitud* (MLE) del parámetro desconocido θ como el valor de θ que maximiza $L(\theta)$. El estimador del parámetro θ se representa: $\hat{\theta}$. Una consecuencia inmediata de la definición del estimador es la siguiente propiedad:

$$L(\theta) \leq L(\hat{\theta}) \quad \text{para todos los posibles valores de } \theta \quad (8.20)$$

En el **caso continuo**, no puede darse una explicación tan intuitiva, ya que la probabilidad de que una variable aleatoria tome un valor determinado es siempre cero. En el caso continuo, si $f_X(x, \theta)$ es la densidad de probabilidad de la distribución con un parámetro desconocido θ , la función de verosimilitud se define de la forma siguiente:

$$L(\theta) = f_X(x_1, \theta) \cdot f_X(x_2, \theta) \cdot \dots \cdot f_X(x_n, \theta) \quad (8.21)$$

El estimador de máxima verosimilitud, $\hat{\theta}$, del parámetro desconocido θ , se define como el valor de θ que maximiza $L(\theta)$ para todos los valores permitidos de θ .

El estimador de máxima verosimilitud es único para los parámetros de la mayoría de las familias de distribuciones, es decir, no hay varios valores del parámetro para los cuales se obtenga el mismo valor de la función verosimilitud y éste sea máximo. Más aun, generalmente los estimadores máximo-verosímiles pueden obtenerse mediante métodos de cálculo, ya que el máximo relativo de la función de verosimilitud obtenido diferenciando $L(\theta)$ con respecto a θ e igualando la derivada a cero es generalmente el máximo absoluto.

Una propiedad importante de los estimadores de máxima verosimilitud es que son *invariantes*. En ocasiones puede emplearse esta propiedad para obtener la expresión matemática de los estimadores. Que un estimador sea invariante significa que si $\hat{\theta}$ es el estimador del parámetro θ , y se define un segundo parámetro ν como una función de θ : $\nu = g(\theta)$, entonces se satisface: $\hat{\nu} = g(\hat{\theta})$.

En los Apéndices B y C, situados al final del libro, se indican los estimadores de máxima verosimilitud de los parámetros de las distribuciones más frecuentemente usadas. A continuación se muestran algunos ejemplos de cálculo del estimador.

Estimador de máxima verosimilitud del parámetro de una distribución geométrica

La función de verosimilitud de n observaciones independientes, x_1, \dots, x_n , distribuidas geom(p) es la siguiente (ver Apéndice C):

$$L(p) = (p \cdot (1-p)^{x_1}) \cdot (p \cdot (1-p)^{x_2}) \cdot \dots \cdot (p \cdot (1-p)^{x_n}) = p^n \cdot (1-p)^{\sum_{i=1}^n x_i} \quad (8.22)$$

Puesto que el valor de p que maximiza $L(p)$ es el mismo que maximiza $\ln(L(p))$, a fin de simplificar los cálculos es preferible resolver el problema de puntos extremos del logaritmo de la expresión anterior. Aplicando la función logaritmo a ambos lados de la igualdad en la ecuación 8.22, se obtiene:

$$\ln(L(p)) = n \cdot \ln(p) + \sum_{i=1}^n x_i \cdot \ln(1-p) \quad (8.23)$$

Para calcular el estimador de p se procede de la manera siguiente:

1. Se calcula la primera derivada de la ecuación 8.23. Se obtiene:

$$\frac{d \ln(L(p))}{dp} = \frac{n}{p} - \frac{\sum_{i=1}^n x_i}{1-p} \quad (8.24)$$

2. Se calcula qué valor del parámetro p hace cero la primera derivada (ecuación 8.24). Se obtiene:

$$\frac{n}{\hat{p}} - \frac{\sum_{i=1}^n x_i}{1-\hat{p}} = 0 \rightarrow \hat{p} = \frac{1}{\frac{\sum_{i=1}^n x_i}{n} + 1} = \frac{1}{\bar{X}(n) + 1} \quad (8.25)$$

3. Se calcula la segunda derivada de la ecuación 8.23 (para ello se deriva la ecuación 8.24) y se examina su signo para el valor del punto crítico obtenido en la ecuación 8.25: \hat{p} . Se obtiene que el signo de la segunda derivada es siempre negativo, con lo cual el punto crítico es un máximo:

$$\frac{d^2 \ln(L(p))}{dp^2} = -\frac{n}{p^2} - \frac{\sum_{i:1}^n x_i}{(1-p)^2} < 0 \quad (8.26)$$

Estimador de máxima verosimilitud del parámetro de escala de la distribución exponencial

Dado un conjunto de n observaciones independientes, la función de verosimilitud del parámetro de escala de la distribución exponencial es la siguiente (ver Anexo B):

$$L(\beta) = \left(\frac{1}{\beta} \cdot e^{-\frac{x_1}{\beta}}\right) \cdot \left(\frac{1}{\beta} \cdot e^{-\frac{x_2}{\beta}}\right) \cdot \dots \cdot \left(\frac{1}{\beta} \cdot e^{-\frac{x_n}{\beta}}\right) = \frac{1}{\beta^n} \cdot e^{-\frac{\sum_{i:1}^n x_i}{\beta}} \quad (8.27)$$

Puesto que el valor de β que maximiza $L(\beta)$ es el mismo que maximiza $\ln(L(\beta))$, con el fin de simplificar los cálculos se resolverá el problema de extremos del logaritmo de la ecuación 8.27. Tomando el logaritmo de ambos lados de la igualdad de la ecuación 8.27 se obtiene:

$$\ln(L(\beta)) = -n \cdot \ln(\beta) - \frac{1}{\beta} \sum_{i:1}^n x_i \quad (8.28)$$

La derivada de la ecuación 8.28 respecto a β es:

$$\frac{d \ln(L(\beta))}{d\beta} = -n \cdot \frac{1}{\beta} + \frac{1}{\beta^2} \sum_{i:1}^n x_i \quad (8.29)$$

Igualando a cero la ecuación 8.29, se obtiene el valor de la única raíz de la ecuación, que coincide con la media de los datos:

$$-n \cdot \frac{1}{\hat{\beta}} + \frac{1}{\hat{\beta}^2} \sum_{i:1}^n x_i = 0 \rightarrow \hat{\beta} = \frac{\sum_{i:1}^n x_i}{n} = \bar{X}(n) \quad (8.30)$$

Para analizar si la raíz de la primera derivada es un máximo, un mínimo o un punto de inflexión de 8.28, en primer lugar es preciso calcular la segunda derivada de la ecuación 8.28 (obtenida derivando la ecuación 8.29):

$$\frac{d^2 \ln(L(\beta))}{d\beta^2} = +n \cdot \frac{1}{\beta^2} - \frac{2}{\beta^3} \sum_{i:1}^n x_i \quad (8.31)$$

A continuación debe evaluarse la ecuación 8.31 para el valor de $\hat{\beta}$ obtenido de la ecuación 8.30:

$$\begin{aligned}
n \cdot \frac{1}{\hat{\beta}^2} - \frac{2}{\hat{\beta}^3} \sum_{i:1}^n x_i &= n \cdot \frac{1}{\left(\frac{\sum_{i:1}^n x_i}{n}\right)^2} - \frac{2}{\left(\frac{\sum_{i:1}^n x_i}{n}\right)^3} \sum_{i:1}^n x_i \\
&= \frac{n^3}{\left(\sum_{i:1}^n x_i\right)^2} - \frac{2 \cdot n^3}{\left(\sum_{i:1}^n x_i\right)^2} \\
&= -\frac{n^3}{\left(\sum_{i:1}^n x_i\right)^2} < 0
\end{aligned} \tag{8.32}$$

El resultado obtenido es un número negativo, con lo cual la raíz de la primera derivada es un máximo. Se trata, por tanto, del estimador de máxima verosimilitud del parámetro.

Estimador de máxima verosimilitud del extremo superior del rango de una distribución uniforme

Se pretende calcular el estimador de máxima verosimilitud del extremo superior del rango de una distribución uniforme. Es decir, dada una distribución uniforme $U(0, b)$, cuyo rango es $[0, b]$, se desea calcular el estimador de b . El cálculo se realiza a partir de n muestras independientes de la distribución: x_1, \dots, x_n .

La función de verosimilitud de un conjunto x_1, \dots, x_n de n observaciones es la siguiente:

$$L(b) = \left(\frac{1}{b}\right)^n \tag{8.33}$$

El valor máximo es esta función corresponde al menor valor posible de b , que es igual a la muestra de mayor valor:

$$\hat{b} = \text{máx} \{x_1, \dots, x_n\} = \text{máx}_{1 \leq i \leq n} x_i \tag{8.34}$$

Estimador de máxima verosimilitud del parámetro de forma y del parámetro de escala de una distribución gamma

Consideremos una distribución gamma (α, β) , con parámetro de forma α y parámetro de escala β . Por definición estos parámetros deben satisfacer (ver Anexo B): $\alpha > 0$ y $\beta > 0$. La función de verosimilitud, calculada a partir de n muestras independientes, x_1, \dots, x_n , es la siguiente:

$$\begin{aligned}
L(\alpha, \beta) &= \frac{x_1^{\alpha-1} \cdot e^{-\frac{x_1}{\beta}}}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \frac{x_2^{\alpha-1} \cdot e^{-\frac{x_2}{\beta}}}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \dots \cdot \frac{x_n^{\alpha-1} \cdot e^{-\frac{x_n}{\beta}}}{\beta^\alpha \cdot \Gamma(\alpha)} \\
&= \frac{\left(\prod_{i:1}^n x_i\right)^{\alpha-1} \cdot e^{-\frac{1}{\beta} \cdot \sum_{i:1}^n x_i}}{\beta^{n \cdot \alpha} \cdot \Gamma^n(\alpha)}
\end{aligned} \tag{8.35}$$

Puesto que la función logaritmo es monótona creciente, los valores de α y de β que maximizan $L(\alpha, \beta)$ también maximizan $\ln(L(\alpha, \beta))$. Los cálculos se simplifican notablemente si, en lugar de calcular el máximo de $L(\alpha, \beta)$ se calcula el máximo de $\ln(L(\alpha, \beta))$. Para ello, se sigue la siguiente secuencia de cálculo:

1. Cálculo del logaritmo de la ecuación 8.35:

$$\ln(L(\alpha, \beta)) = (\alpha - 1) \cdot \sum_{i:1}^n \ln(x_i) - \frac{1}{\beta} \cdot \sum_{i:1}^n x_i - n \cdot \alpha \cdot \ln(\beta) - n \cdot \ln(\Gamma(\alpha)) \quad (8.36)$$

2. Cálculo de la derivada parcial de $\ln(L(\alpha, \beta))$ respecto a α y β :

$$\frac{\partial \ln(L(\alpha, \beta))}{\partial \alpha} = \sum_{i:1}^n \ln(x_i) - n \cdot \ln(\beta) - n \cdot \frac{1}{\Gamma(\alpha)} \cdot \frac{d\Gamma(\alpha)}{d\alpha} \quad (8.37)$$

$$\frac{\partial \ln(L(\alpha, \beta))}{\partial \beta} = \frac{1}{\beta^2} \cdot \sum_{i:1}^n x_i - \frac{n \cdot \alpha}{\beta} \quad (8.38)$$

3. Cálculo de los valores de α y β que hacen cero simultáneamente las dos ecuaciones anteriores. Estos valores se obtienen resolviendo el sistema de ecuaciones siguiente:

$$\begin{cases} \ln(\beta) + \frac{1}{\Gamma(\alpha)} \cdot \frac{d\Gamma(\alpha)}{d\alpha} = \frac{\sum_{i:1}^n \ln(x_i)}{n} \\ \alpha \cdot \beta = \frac{\sum_{i:1}^n x_i}{n} \end{cases} \quad (8.39)$$

Este sistema de ecuaciones debe resolverse empleando métodos numéricos. El valor de los parámetros que lo resuelven son los estimadores de máxima verosimilitud, ya que puede demostrarse que son un máximo de la función $\ln(L(\alpha, \beta))$. La función $\frac{1}{\Gamma(\alpha)} \cdot \frac{d\Gamma(\alpha)}{d\alpha}$ se denomina *función digamma*.

Intervalo de confianza de los estimadores de máxima verosimilitud

Los estimadores de máxima verosimilitud pueden ser estimadores sesgados, es decir, el valor esperado de $\hat{\theta}$, $E\{\hat{\theta}\}$, puede no coincidir con el valor verdadero (desconocido) del parámetro θ .

Sin embargo, la distribución asintótica del estimador $\hat{\theta}$ cuando el tamaño de la muestra tiende a infinito (es decir, $n \rightarrow \infty$) es una distribución normal cuya media es igual a θ y su varianza igual a $\delta(\theta) = \frac{-n}{E\left\{\frac{d^2 \ln(L(\theta))}{d\theta^2}\right\}}$.

Equivalentemente:

$$\frac{\hat{\theta} - \theta}{\sqrt{\frac{\delta(\hat{\theta})}{n}}} \text{ en el límite } n \rightarrow \infty \text{ está distribuido } N(0, 1) \quad (8.40)$$

Por consiguiente, cuando n es suficientemente grande, el siguiente es un intervalo con un $100 \cdot (1 - \alpha) \%$ de confianza para el parámetro θ :

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\delta(\hat{\theta})}{n}} \quad (8.41)$$

Es importante pensar en los parámetros de entrada de la simulación en términos de sus intervalos de confianza. No debe olvidarse que los estimadores de estos parámetros son variables aleatorias, cuya distribución depende del tamaño de la muestra que se ha usado para calcularlos. Cuanto mayor es el tamaño de la muestra, menor es la varianza de la distribución del estimador y, por tanto, menor es la longitud de su intervalo de confianza.

En ocasiones, la respuesta del modelo puede ser muy sensible frente a variaciones en el valor de determinado parámetro de entrada. Esta posibilidad debe ser analizada para cada modelo en concreto. Para ello, debe simularse el sistema no sólo para el valor estimado de los parámetros de entrada, sino también para los valores extremos de sus intervalos de confianza (frecuentemente del 90, 95 ó 99 %) y comparar las respuestas obtenidas.

Si se observara que la dependencia de la respuesta del sistema respecto a un parámetro, cuando éste varía dentro de su intervalo de confianza, es relevante para las conclusiones del estudio de simulación, debe estimarse el parámetro con mayor precisión a fin de reducir la longitud de su intervalo de confianza.

8.6. TAREA III: Medida de la bondad del ajuste

Una vez estimados los parámetros de la distribución a partir de los datos experimentales x_1, \dots, x_n , cabe preguntarse en qué medida los datos experimentales responden a la distribución ajustada, $F_X(x)$. Se trata de responder a la pregunta: ¿cabría haber obtenido los datos experimentales muestreando la distribución $\hat{F}_X(x)$ o, por el contrario, existen grandes discrepancias entre la distribución ajustada y los datos experimentales?

Para intentar responder a esta pregunta se usan técnicas gráficas y tests estadísticos. Los *tests de ajuste* pretenden contrastar la hipótesis siguiente:

$$H_0: \text{Las observaciones } x_1, \dots, x_n \text{ están distribuidas } \hat{F}_X(x).$$

Hay que tener en cuenta que la mayoría de los tests no son demasiado potentes: cuando se dispone de pocas observaciones son poco sensibles a discrepancias entre los datos y la distribución ajustada¹¹ y, sin embargo, cuando se dispone de muchas observaciones una pequeña discrepancia puede hacer que se rechace el ajuste.

Si bien la bibliografía sobre tests de ajuste es muy extensa, en este tema únicamente se describen dos de ellos: el *test chi-cuadrado* y el *test de Kolmogorov-Smirnov*.

La decisión acerca de cuál de estos dos tests usar depende de la naturaleza de la distribución y del tamaño de la muestra x_1, \dots, x_n :

- El test de Kolmogorov-Smirnov es sólo válido para distribuciones continuas, mientras que el test chi-cuadrado es aplicable tanto a distribuciones discretas como continuas.
- El test chi-cuadrado requiere la clasificación de los datos de la muestra en clases. Dado que se recomienda que haya al menos 5 observaciones por clase y conviene tener un número razonablemente grande de clases, el test chi-cuadrado no es aplicable a muestras pequeñas¹². Por el contrario, el test de Kolmogorov-Smirnov es aplicable a cualquier tamaño de muestra.

Técnicas gráficas

Las técnicas gráficas son una herramienta muy potente de análisis. Comúnmente se usan el mismo tipo de representaciones que para seleccionar la familia de distribuciones (Tarea I), es decir, histogramas y gráficas Q-Q.

Se muestra un ejemplo en la Figura 8.6. Se han tomado 100 muestras de un proceso de llegada (tiempos entre llegadas sucesivas). Apoyándose en razonamientos teóricos y en

¹¹Estos tests tienen generalmente una probabilidad baja de rechazar un ajuste incorrecto. Como resultado de ello, se da la situación de ajustar diferentes distribuciones a los datos y que estos tests no rechacen ninguno de los ajustes. Según se recomienda en (Pedgen et al. 1995), el hecho de que un test no rechace un ajuste no debe considerarse como una fuerte evidencia de que la distribución escogida sea un buen ajuste: en la práctica conviene complementar el uso de los tests con el análisis visual del ajuste usando técnicas gráficas.

¹²En (Pedgen et al. 1995) se recomienda no usar el test chi-cuadrado para tamaños de muestra inferiores a 100, si bien se señala que algunos autores indican que han obtenido buenos resultados con tamaños de muestra tan pequeños como 30.

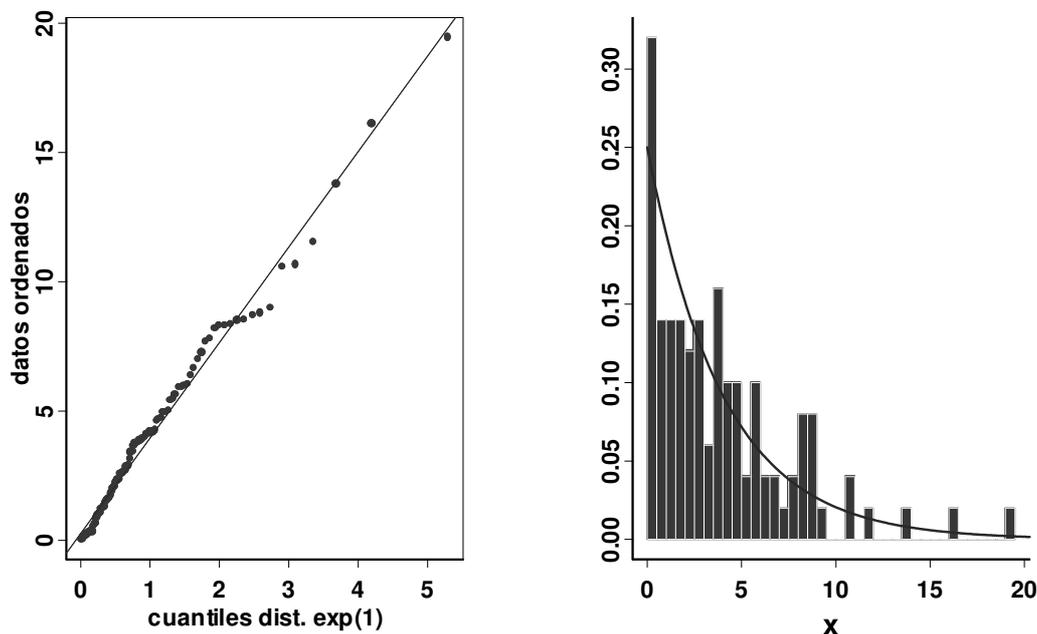


Figura 8.6: Ejemplo. Gráfica Q-Q (izqda). Histograma y densidad (drcha).

la gráfica Q-Q (ver gráfica Q-Q, en la parte izquierda de la Figura 8.6), se ha realizado la hipótesis de que los datos están distribuidos exponencialmente. El estimador del parámetro de la distribución se calcula de los datos experimentales: $\hat{\beta} = 4$. A continuación se compara la distribución ajustada con los datos experimentales. Para ello, se representa el histograma de los datos experimentales escalado como una densidad de probabilidad (la suma del área de las barras vale uno) y la densidad de probabilidad de la distribución ajustada: $\text{expo}(4)$. Este gráfico se muestra en la parte derecha de la Figura 8.6.

Test χ^2

El test χ^2 (chi-cuadrado) puede considerarse un método formal para:

- En el caso continuo, comparar el histograma (escalado como densidad) de un conjunto de datos experimentales independientes, x_1, \dots, x_n , con la densidad de probabilidad de determinada distribución $\hat{f}_X(x)$.
- En el caso discreto, comparar la probabilidad experimental de cada valor de la variable aleatoria (obtenida de un conjunto de datos experimentales independientes x_1, \dots, x_n) con la correspondiente probabilidad de determinada distribución \hat{p}_X .

La hipótesis que se desea contrastar es:

H_0 : Las observaciones x_1, \dots, x_n y la distribución están igualmente distribuidas.

El test es aplicable a cualquier distribución. La aplicación del test χ^2 consta de los siguientes pasos:

1. *Definición de las clases.* En primer lugar se divide el rango de la distribución en k intervalos adyacentes disjuntos (denominados clases o categorías):

$$[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$$

Puede suceder que $a_0 = -\infty$ y/o que $a_k = \infty$, con lo cual el primer y último intervalos serían $(-\infty, a_1)$ y $[a_{k-1}, \infty)$ respectivamente.

2. *Cálculo de la frecuencia observada.* Se define N_j como el número de observaciones del conjunto x_1, \dots, x_n que están incluidas en el intervalo j -ésimo: $[a_{j-1}, a_j)$. Puesto que los intervalos recubren el rango de la distribución, se satisface: $\sum_{j:1}^k N_j = n$.
3. *Cálculo de la frecuencia esperada.* De acuerdo con la distribución de probabilidad ($\hat{f}_X(x)$ en el caso continuo y \hat{p}_X en el discreto), la probabilidad de que una observación caiga en el intervalo j -ésimo es:

$$p_j = \text{Prob} \{a_{j-1} < X \leq a_j\} \quad (8.42)$$

Puede calcularse de la forma siguiente:

$$\begin{aligned} p_j &= \int_{a_{j-1}}^{a_j} \hat{f}_X(x) \cdot dx && \text{Caso continuo} \\ p_j &= \sum_{\{i: a_{j-1} \leq x_i < a_j\}} \hat{p}_X(x_i) && \text{Caso discreto} \end{aligned} \quad (8.43)$$

4. *Cálculo del estadístico.* Se define el estadístico siguiente:

$$\chi^2 = \sum_{j:1}^k \frac{(N_j - n \cdot p_j)^2}{n \cdot p_j} \quad (8.44)$$

Su significado es muy intuitivo:

$n \cdot p_j$ *Frecuencia esperada.* Es el número de observaciones que caen en el intervalo j -ésimo de acuerdo a la distribución.

N_j *Frecuencia observada.* Es el número de observaciones experimentales que han caído en el intervalo j -ésimo.

Cabe esperar que cuanto mejor sea el ajuste, menor sea el valor del estadístico, χ^2 . Por consiguiente, se rechaza la hipótesis nula cuando χ^2 sea "demasiado grande".

5. *Contraste de la hipótesis.* La distribución del estadístico χ^2 depende de si se han usado o no los datos experimentales x_1, \dots, x_n para ajustar uno o más parámetros de la distribución. A continuación se considera cada una de estas dos opciones por separado:
 - No se han usado los datos experimentales x_1, \dots, x_n para ajustar ningún parámetro de la distribución. En este caso, si H_0 es cierta, entonces cuando $n \rightarrow \infty$ el estadístico χ^2 está distribuido chi-cuadrado con $k-1$ grados de libertad. Este hecho permite establecer, con una confianza del $100 \cdot (1 - \alpha) \%$, el criterio del test:

$$\chi^2 \begin{cases} > \chi_{k-1, 1-\alpha}^2 & \rightarrow \text{se rechaza } H_0 \\ \leq \chi_{k-1, 1-\alpha}^2 & \rightarrow \text{no hay evidencia para rechazar } H_0 \end{cases} \quad (8.45)$$

- Uno o varios parámetros de la distribución se han estimado a partir de los datos experimentales, x_1, \dots, x_n , y además el valor de estos parámetros es su estimador de máxima verosimilitud. Sea m el número de parámetros de la distribución estimados usando los datos x_1, \dots, x_n .

En este caso, si la hipótesis nula H_0 es cierta, entonces cuando $n \rightarrow \infty$ el estadístico χ^2 converge a una distribución que está comprendida entre χ_{k-1}^2 (chi-cuadrado con $k-1$ grados de libertad) y χ_{k-m-1}^2 (chi-cuadrado con $k-m-1$ grados de libertad).

Sea $\chi_{\alpha-1}^2$ el punto crítico $1 - \alpha$ de la distribución del estadístico χ^2 . Puesto que se desconoce la distribución asintótica del estadístico, no es posible calcular este punto crítico. No obstante, si es posible acotarlo de la forma siguiente (al aumentar los grados de libertad, el valor de los puntos críticos crece. Ver Figura 8.7):

$$\chi_{k-m-1, \alpha-1}^2 \leq \chi_{\alpha-1}^2 \leq \chi_{k-1, \alpha-1}^2 \quad (8.46)$$

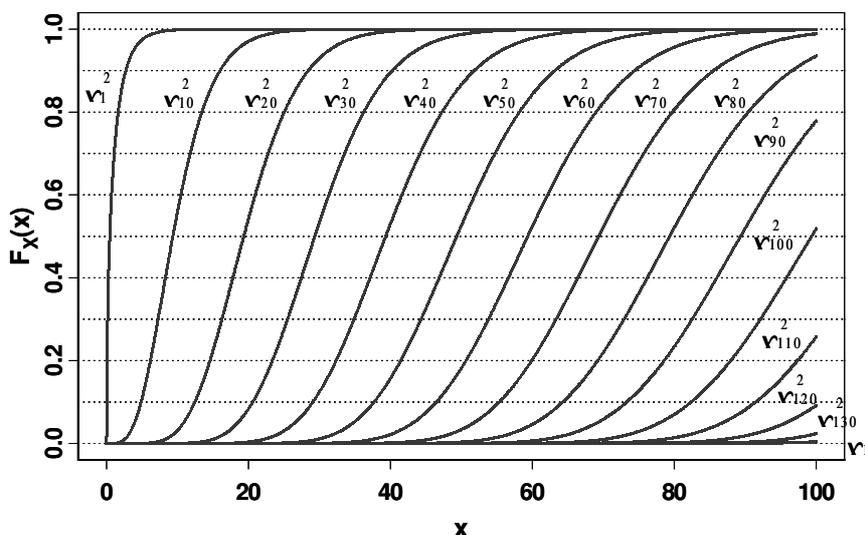


Figura 8.7: Probabilidad acumulada de algunas distribuciones chi-cuadrado.

Por tanto, adoptando la postura conservadora¹³, se llega al mismo criterio que en el caso anterior (cuando no se usan los datos para estimar ningún parámetro de la distribución):

$$\chi^2 \begin{cases} > \chi_{k-1,1-\alpha}^2 & \rightarrow \text{se rechaza } H_0 \\ \leq \chi_{k-1,1-\alpha}^2 & \rightarrow \text{no hay evidencia para rechazar } H_0 \end{cases} \quad (8.47)$$

Posiblemente el aspecto más problemático de este test es la elección de las clases (intervalos en que se divide el rango de la distribución). No existe un método general que garantice buenos resultados para cualquier distribución ajustada y para cualquier tamaño de muestra. Incluso en ocasiones, pueden obtenerse conclusiones diferentes, a partir del mismo conjunto de observaciones, dependiendo de la elección de las clases. Esta es la mayor desventaja del test.

Existen, sin embargo, algunas recomendaciones que son generalmente aceptadas¹⁴, si bien no proporcionan una respuesta completa al problema de la elección de las clases. Dichas recomendaciones son las siguientes:

- En la medida de lo posible, es conveniente escoger las clases de modo que se satisfaga: $p_1 = p_2 = \dots = p_k$
- Escoger los intervalos de modo que las frecuencias esperadas $n \cdot p_j$ no sean demasiado pequeñas, por ejemplo: $n \cdot p_j \geq 5$ para $j : 1, \dots, k$.
- El número de clases afecta a la potencia del test, si bien el valor de k que conduce a la máxima potencia depende de la forma de la distribución. En general, suele aconsejarse no escoger k mayor de 30 ó 40.

¹³En (Law & Kelton 2000) se realiza esta recomendación: rechazar la hipótesis nula H_0 si y sólo si se satisface $\chi^2 > \chi_{k-1,1-\alpha}^2$. Con este criterio, la probabilidad de estar cometiendo un error de Tipo I (rechazar H_0 cuando es verdadera) es al menos tan pequeña como la probabilidad deseada: α . Sin embargo, esta elección reduce la potencia del test (probabilidad de rechazar H_0 cuando es falsa).

En otros textos, como por ejemplo (Pedgen et al. 1995) se recomienda rechazar H_0 si y sólo si se satisface $\chi^2 > \chi_{k-m-1,1-\alpha}^2$. En la práctica, m no es mayor que 2. Si el número de clases k es grande, entonces la diferencia entre $\chi_{k-1,1-\alpha}^2$ y $\chi_{k-m-1,1-\alpha}^2$ no es demasiado grande.

¹⁴La razón de cada una de estas recomendaciones se discute en (Law & Kelton 2000).

Test de Kolmogorov-Smirnov

La finalidad del test es realizar la comparación de dos funciones de probabilidad acumuladas: la distribución empírica obtenida de los datos experimentales x_1, \dots, x_n , es decir: $F_X(x) = \frac{\text{número de } x_i \text{'s } \leq x}{n}$, con determinada probabilidad acumulada, $\hat{F}(x)$. La hipótesis que se desea contrastar es la siguiente:

$$H_0: \text{Las distribuciones } F(x) \text{ y } \hat{F}(x) \text{ son iguales.}$$

Dado que no es preciso agrupar los datos experimentales en clases, este test no presenta el problema de la selección de las clases.

Se emplea el siguiente estadístico para contrastar la hipótesis nula:

$$D = \sup_x \left\{ \left| F_X(x) - \hat{F}(x) \right| \right\} \quad (8.48)$$

donde \sup representa el *supremo* del conjunto. El *supremo* de un conjunto es el valor más pequeño tal que cumple que es mayor o igual que todos los elementos del conjunto. Si el supremo pertenece al conjunto, entonces coincide con el máximo.

La distribución del estadístico D depende de:

- la distribución $\hat{F}(x)$,
- el número de datos experimentales, n , y
- de si se han usado los datos experimentales, x_1, \dots, x_n para ajustar algún parámetro de la distribución.

El test con nivel α consiste en comprobar si el valor del estadístico excede el valor del punto crítico de su distribución correspondiente al % de confianza elegido.

8.7. Selección de la distribución en ausencia de datos

En algunos estudios de simulación, no se dispone de datos para estimar la distribución de probabilidad de las entradas aleatorias. A continuación, se discuten algunas recomendaciones que pueden ayudar a seleccionar la distribución en ausencia de datos¹⁵.

- Si el proceso tiene muy poca variabilidad y se conoce su valor medio, puede sustituirse en el modelo la entrada aleatoria por un parámetro determinista cuyo valor sea el valor medio.
- Cuando un proceso estocástico de llegada cumple determinadas condiciones (que se discutirán en la próxima sección), suele usarse una distribución exponencial para generar los intervalos de tiempo entre llegadas sucesivas. No obstante, las distribuciones exponenciales deben usarse con cautela, ya que tienen una variabilidad muy grande. Por ejemplo, normalmente una distribución exponencial no resulta adecuada para modelar tiempos de servicio¹⁶, ya que frecuentemente la duración de los mismos no tiene una variabilidad tan grande como representa la distribución exponencial.
- Si se conoce la media y se sabe que la variabilidad del proceso es cierto porcentaje de la misma, puede usarse en primera aproximación una distribución uniforme o triangular simétrica. Por ejemplo, si se sabe que la media es 100 y que la variabilidad está entorno al 30 %, puede usarse $U(70, 130)$ o bien $\text{triang}(70, 130, 100)$.

Debe tenerse en cuenta que los resultados obtenidos *dependen fuertemente de la distribución seleccionada*.

¹⁵Estas y otras recomendaciones útiles pueden encontrarse en los textos (Pedgen et al. 1995) y (Law & Kelton 2000).

¹⁶Aunque la distribución exponencial se usa en algunos modelos analíticos para representar tiempos de servicio, el motivo de ello es simplificar los cálculos matemáticos y no el hecho de que represente de manera realista los tiempos de servicio.

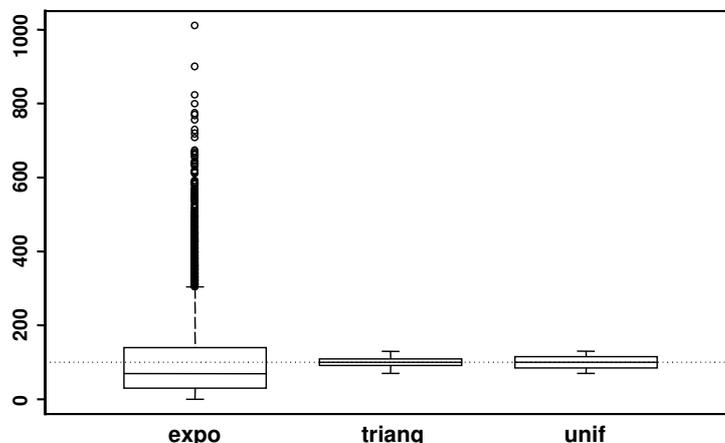


Figura 8.8: Boxplots de las muestras.

Por ejemplo, en la Figura 8.8 se muestran los boxplots de 10.000 muestras extraídas de cada una de las siguientes tres distribuciones: expo (100), U (70, 130) y triang (70, 130, 100). La media, mediana, desviación estándar, el máximo y el mínimo de cada una de estas tres muestras son las siguientes:

| | Media | Mediana | Desviación Estándar | Mínimo | Máximo |
|-----------------------|----------|----------|---------------------|---------|-----------|
| expo (100) | 101.1192 | 69.1468 | 101.0142 | 0.0069 | 1010.9270 |
| triang (70, 130, 100) | 100.0656 | 100.1020 | 12.2538 | 70.2129 | 129.4429 |
| U (70, 130) | 99.9758 | 100.3433 | 17.2262 | 70.0022 | 129.9875 |

Como puede observarse, la distribución exponencial tiene una variabilidad mucho mayor que la triangular y la uniforme.

- En algunos casos se conocen los valores máximo y mínimo del proceso. Dependiendo de la naturaleza del proceso, puede usarse una distribución triangular con la media en el punto central del rango.

Cuando sólo se conocen los extremos del rango, algunos autores recomiendan usar una distribución uniforme. Sin embargo, es difícil imaginar una situación real en la cual el tiempo de servicio, el intervalo entre llegadas o entre averías de una máquina, etc., esté uniformemente distribuido.

En algunos casos está justificado usar una distribución normal, con media en el punto medio del rango y desviación estándar igual a un cuarto del rango (es decir, rango = $4 \cdot \sigma$). Como recomendación práctica, se emplea un cuarto del rango en lugar de un sexto debido a que parece demostrado que comúnmente se infraestima el rango de valores posibles olvidando o ignorando los extremos. El problema de usar una distribución normal es que su rango es $(-\infty, \infty)$. En la mayoría de las situaciones un rango infinito no tiene sentido, ya que puede dar lugar a valores no realistas. Una forma de resolver el problema es truncar la distribución para evitar la generación de valores demasiado grandes o demasiado pequeños. Por ejemplo, puede truncarse al rango $\mu \pm 4 \cdot \sigma$.

El resultado obtenido depende de la distribución escogida, si bien la dependencia no es tan pronunciada como en el caso anterior. Considérese, por ejemplo, que el rango del proceso es: mínimo=80, máximo=120. Se modela mediante tres distribuciones: U (80, 120), triang (80, 120, 100) y N ($\mu = \frac{120+80}{2} = 100, \sigma = \frac{120-80}{4} = 10$). Se han generado 10.000 mues-

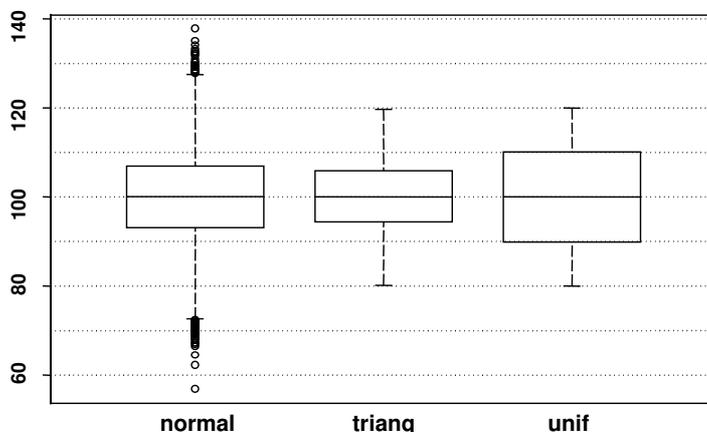


Figura 8.9: Boxplots de las muestras.

tras de cada distribución. Los resultados se muestran en la Figura 8.9. Los estadísticos de las observaciones se muestran a continuación:

| | Media | Mediana | Desviación Estándar | Mínimo | Máximo |
|-----------------------|----------|----------|---------------------|----------|---------|
| N (100, 10) | 100.0424 | 100.1015 | 10.1002 | 137.7396 | 56.8264 |
| triang (80, 120, 100) | 100.1135 | 100.0185 | 8.0703 | 119.6908 | 80.1268 |
| U (80, 120) | 99.9957 | 100.0066 | 11.6050 | 119.9972 | 80.0082 |

- Cuando se conoce el mínimo, el máximo y el modo (valor más probable del proceso, es decir, el máximo de la densidad de probabilidad), posiblemente las mejores opciones son usar la distribución triangular, la beta o la normal.

La distribución triangular es la más cómoda de usar, ya que el máximo, mínimo y el modo son directamente los parámetros de la distribución.

Otra posibilidad es la distribución beta: se usa frecuentemente para modelar el tiempo requerido para completar una tarea. Esta distribución ofrece mucha flexibilidad de modelado, debido a la gran variedad de formas que la distribución beta puede asumir. Para escoger los parámetros α_1 y α_2 de la distribución, debe tenerse en cuenta que:

- El rango de la distribución beta (α_1, α_2) es $[0, 1]$, con lo cual en general es preciso reescalar y desplazar la distribución. Se realiza de la forma siguiente:

$$\begin{array}{l} X \sim \text{beta}(\alpha_1, \alpha_2), \text{ rango}(X) = [0, 1] \\ Y = a + (b - a) \cdot X \end{array} \quad \left| \rightarrow \quad \begin{array}{l} Y \sim \text{beta}(\alpha_1, \alpha_2) \\ \text{rango}(Y) = [a, b] \end{array} \quad (8.49)$$

- La media (μ) y el modo (c) de una distribución beta (α_1, α_2) con rango $[a, b]$ y con valores de los parámetros $\alpha_1 > 1$ y $\alpha_2 > 1$, son los siguientes:

$$\mu = a + \frac{\alpha_1 \cdot (b - a)}{\alpha_1 + \alpha_2} \quad (8.50)$$

$$c = a + \frac{(\alpha_1 - 1) \cdot (b - a)}{\alpha_1 + \alpha_2 - 2} \quad (8.51)$$

Si a partir del conocimiento que se tiene sobre el proceso es posible estimar la media y el modo de la distribución, entonces resolviendo las ecuaciones 8.50 y 8.51 puede calcularse la correspondiente estimación de los parámetros α_1 y α_2 :

$$\alpha_1 = \frac{(\mu - a) \cdot (2 \cdot c - a - b)}{(c - \mu) \cdot (b - a)} \quad (8.52)$$

$$\alpha_2 = \frac{b - \mu}{\mu - a} \cdot \alpha_1 \quad (8.53)$$

En función del valor que se asigne a la media y al modo, la distribución beta estará sesgada hacia la derecha (caso $\mu > c$) o hacia la izquierda (caso $\mu < c$). Puesto que se ha partido del supuesto de que se conoce el máximo, el mínimo y el modo, queda por estimar la media del proceso. Una posible forma de hacerlo es la siguiente (Pedgen et al. 1995):

$$\tilde{\mu} = \frac{a + 4 \cdot c + b}{6} \quad (8.54)$$

8.8. Procesos estocásticos de Poisson

Procesos estocásticos de llegada

Una proporción importante de las entradas a la simulación lo constituyen los procesos estocásticos de llegada de las entidades al sistema.

Supóngase la siguiente secuencia de instantes aleatorios: $0 = t_0 \leq t_1 \leq t_2 \leq \dots$, tal que en cada uno de estos instantes de tiempo sucede un determinado evento (por ejemplo, la llegada de una entidad al sistema), y de modo que todos los instantes de tiempo $\{t_i\}$ son muestras de una determinada distribución de probabilidad (es decir, son observaciones de variables aleatorias idénticamente distribuidas).

Asimismo, los intervalos de tiempo entre llegadas sucesivas, definidos como:

$$A_i = t_i - t_{i-1} \text{ con } i : 1, 2, \dots \quad (8.55)$$

son observaciones independientes de variables aleatorias idénticamente distribuidas.

Dado que en cada instante t_i se produce un único evento, el número total de eventos, representado $N(t)$, que se han producido hasta el instante t (inclusive), puede calcularse de la forma siguiente:

$$N(t) = \text{máx} \{i \text{ tal que } t_i \leq t\} \quad (8.56)$$

En aquellos casos en que los eventos son llegadas de entidades al sistema, se dice que el proceso estocástico $\{N(t), t \geq 0\}$ es un *proceso estocástico de llegadas*.

Procesos de Poisson

El proceso estocástico de llegadas $\{N(t), t \geq 0\}$ se dice que es un *proceso de Poisson* cuando cumple las tres condiciones siguientes:

- *Condición 1.* Las entidades llegan al sistema de una en una, es decir, en cada instante $\{t_i\}$ se produce la llegada de una única entidad. Esta condición no se cumple si las entidades llegan en grupos de diferente tamaño.
- *Condición 2.* El número de llegadas que se produce en el intervalo de tiempo $(t, t + s]$, es decir $N(t + s) - N(t)$, es independiente de cómo se haya desarrollado el proceso con anterioridad al instante de tiempo t . Esta condición no se cumple, por ejemplo, si un cliente deja de enviar solicitudes de compra (entidades que llegan al sistema) cuando el sistema tiene una larga cola de solicitudes pendientes de atender.

- *Condition 3.* El proceso es *estacionario*. La distribución del número de llegadas $N(t+s) - N(t)$ que se producen en el intervalo $(t, t+s]$ es independiente del instante de tiempo t en que comience el intervalo. Esta condición debe satisfacerse para cualquier valor de t y para cualquier longitud del intervalo, s . Esto implica que la frecuencia de llegadas es siempre constante, cosa que en muchos sistemas reales no se cumple (la frecuencia depende de la hora del día, la época del año, o de otros factores).

A continuación se van a enunciar dos teoremas y una propiedad, relativos a los procesos de Poisson, que resultan especialmente útiles.

Teorema 8.8.1. Si $\{N(t), t \geq 0\}$ es un proceso de Poisson, entonces el número de llegadas en cualquier intervalo de tiempo de longitud s es una variable aleatoria de Poisson¹⁷ con parámetro $\lambda \cdot s$, donde λ es un número real positivo. Se satisface:

$$\text{Prob}\{(N(t+s) - N(t)) = k\} = \frac{e^{-\lambda \cdot s} \cdot (\lambda \cdot s)^k}{k!} \text{ para } k = 0, 1, 2, \dots \text{ y } t, s \geq 0 \quad (8.57)$$

□

Se cumple: $E\{N(s)\} = \lambda \cdot s$. En particular: $E\{N(1)\} = \lambda$. Por consiguiente, λ es el número esperado de llegadas en cualquier intervalo de tiempo de longitud unidad. El parámetro λ se denomina *frecuencia de llegada* del proceso, y se mide en entidades por unidad de tiempo.

Teorema 8.8.2. Si $\{N(t), t \geq 0\}$ es un proceso de Poisson con frecuencia λ , entonces sus correspondientes intervalos entre llegadas A_1, A_2, \dots son variables aleatorias independientes distribuidas exponencialmente de media $\beta = \frac{1}{\lambda}$.

Lo inverso también es cierto. Si los intervalos de tiempo, A_1, A_2, \dots , entre llegadas de un proceso de llegadas $\{N(t), t \geq 0\}$ son variables aleatorias independientes, distribuidas exponencialmente con media $\beta = \frac{1}{\lambda}$, entonces $\{N(t), t \geq 0\}$ es un proceso de Poisson con frecuencia λ . □

Propiedad 8.8.1. Suponga que se observa un proceso de Poisson durante un intervalo de tiempo $[0, T]$, donde la longitud T del intervalo se fija antes de iniciar las observaciones. En diferentes instantes de tiempo a lo largo del intervalo, t_1, t_2, \dots, t_n , se van produciendo los eventos. Los instantes de tiempo satisfacen: $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq T$. Pues bien, los valores t_1, t_2, \dots, t_n son indistinguibles, desde el punto de vista estadístico de n números que han sido obtenidos muestreando independientemente una distribución $U(0, T)$ y posteriormente ordenados en orden creciente. □

Esta propiedad proporciona un procedimiento para contrastar la hipótesis de que cierta secuencia observada de eventos está generada por un proceso de Poisson. El procedimiento consiste en contrastar si los instantes de los eventos t_1, t_2, \dots, t_n parecen ser observaciones independientes muestreadas de la distribución $U(0, T)$.

Procesos de Poisson no estacionarios

En muchos sistemas reales, la frecuencia de llegada de las entidades al sistema es una función del tiempo, $\lambda(t)$, con lo cual no se satisface la Condición 3. En este caso, los intervalos entre llegadas A_1, A_2, \dots no están idénticamente distribuidos.

Aquellos procesos estocásticos de llegada, $\{N(t), t \geq 0\}$, que satisfacen las Condiciones 1 y 2, pero no la 3, puesto que la frecuencia varía con el tiempo, se denominan *procesos de Poisson no estacionarios*.

En los procesos de Poisson no estacionarios, el valor esperado del número de eventos que se producen en el intervalo $[0, t]$ es función del valor de t .

La *frecuencia de llegada*, $\lambda(t)$ puede definirse formalmente como la derivada temporal del valor esperado del número de llegadas que se producen en el intervalo $[0, t]$:

¹⁷Este teorema explica el origen del nombre "procesos de Poisson".

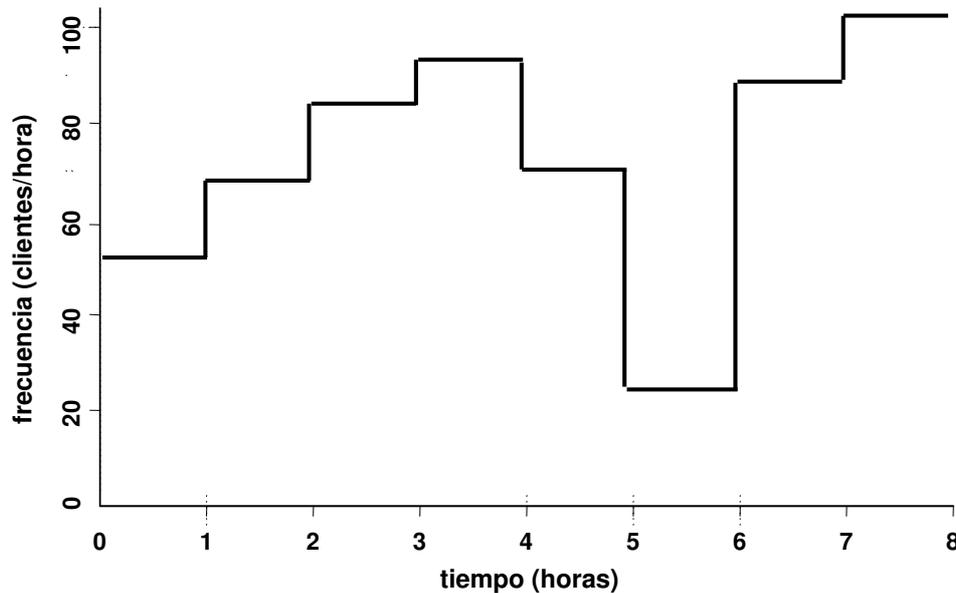


Figura 8.10: Ejemplo de estimación de la evolución temporal de la frecuencia.

$$\lambda(t) = \frac{d\Lambda(t)}{dt} \text{ donde } \Lambda(t) = E\{N(t)\} \quad (8.58)$$

Teorema 8.8.3. Si $\{N(t), t \geq 0\}$ es un proceso de Poisson no estacionario, en el cual $\Lambda(t) = E\{N(t)\}$ es una función continua del tiempo, entonces se satisface:

$$\text{Prob}\{(N(t+s) - N(t)) = k\} = \frac{e^{-b(t,s)} \cdot (b(t,s))^k}{k!} \text{ para } k = 0, 1, 2, \dots \text{ y } t, s \geq 0 \quad (8.59)$$

donde: $b(t, s) = \Lambda(t+s) - \Lambda(t)$ □

Este teorema significa que en un proceso de Poisson no estacionario, el número de llegadas que se producen en el intervalo $(t, t+s]$ es una variable aleatoria de Poisson cuyo parámetro, $b(t, s)$, depende de la longitud del intervalo y del tiempo.

En la práctica, la frecuencia de llegada de las entidades, $\lambda(t)$, puede estimarse de la manera siguiente.

Considérese, por ejemplo, un servicio de atención al cliente que funciona 8 horas al día. La llegada de clientes se modela como un proceso de Poisson no estacionario. Para estimar la evolución de la frecuencia de llegada a lo largo de la jornada laboral, se miden los instantes de llegada de los clientes durante m días. El análisis de los datos se realiza dividiendo la jornada laboral en k intervalos y calculando la media de las llegadas que se producen en cada intervalo en los m días. Dividiendo el número medio de llegadas de cada intervalo por la longitud del intervalo se obtiene una estimación de la frecuencia de llegada en el mismo.

En la Figura 8.10 se muestra la representación de la frecuencia de este proceso calculada para $k = 8$.

Procesos de Poisson compuestos

En algunos sistemas las llegadas se producen en grupos compuestos de un número aleatorio de entidades, con lo que no se cumple la Condición 1. En algunos de estos casos, el proceso de llegada puede modelarse de la siguiente manera:

- Si los intervalos entre llegadas están distribuidos aproximadamente de manera exponencial y son independientes, entonces los instantes de llegada de los grupos de entidades se modelan como un proceso de Poisson: $\{N(t), t \geq 0\}$. En este caso, $N(t)$ representa el número de *grupos* de entidades que han llegado durante el intervalo $[0, t]$.
- En algunos casos, el tamaño de los grupos puede considerarse que son observaciones independientes de cierta distribución discreta de probabilidad. El número de entidades que compone el grupo que llega en el instante t_i se representa: B_i .

De este modo, el número de entidades, $X(t)$ que han llegado en el intervalo $[0, t]$, se calcula sumando las entidades de todos los grupos que han llegado en dicho intervalo de tiempo:

$$X(t) = \sum_{i:1}^{N(t)} B_i \text{ para } t \geq 0 \quad (8.60)$$

Se dice que el proceso estocástico $\{X(t), t \geq 0\}$ es un *proceso de Poisson compuesto* si se cumplen las siguientes condiciones:

- Las variables aleatorias B_i son independientes entre sí y están idénticamente distribuidas.
- Las variables aleatorias B_i son independientes del proceso de llegada $\{N(t), t \geq 0\}$.
- EL proceso $\{N(t), t \geq 0\}$ es de Poisson.

Modelado usando Arena

El módulo “Create” permite el modelado de los procesos de llegada de las entidades al sistema. El tiempo entre llegadas puede ser constante, o estar distribuido de acuerdo a cualquier expresión basada en las distribuciones de probabilidad que soporta Arena. Igualmente, el número de entidades por llegada (“Entities per Arrival”) puede ser constante o estar distribuido aleatoriamente. Por consiguiente, los procesos de llegada de Poisson son un caso particular dentro de los procesos de llegada que permite modelar Arena.

En los *procesos de Poisson estacionarios*, el intervalo de tiempo entre llegadas sucesivas está distribuido exponencialmente, con media constante. Para particularizar un módulo “Create” a este tipo de proceso, puede seleccionarse “Random (Expo)” como tipo (“Type”) del tiempo entre llegadas (“Time Between Arrivals”).

Para modelar *procesos de Poisson no estacionarios* debe seleccionarse, en el módulo “Create”, “Schedule” en el tipo (“Type”) de tiempo entre llegadas (“Time Between Arrivals”). A continuación, debe especificarse la evolución temporal de la frecuencia en el módulo de datos “Schedule”. Se realiza de forma equivalente a la planificación de la capacidad de los recursos, con la única diferencia de que en la columna “Type” de la fila correspondiente debe seleccionarse “Arrival” en lugar de “Capacity”.

Cada par de valores “Value - Duration” de la planificación indica la frecuencia, expresada en *número de llegadas por hora*, que se produce durante el periodo de tiempo indicado. La duración del periodo de tiempo se expresa en las unidades indicadas en la casilla “Time Units”.

Ejemplo 8.8.1. *Considérese la planificación siguiente:*

Name : Llegada de clientes

Type : Arrival

Time Units: days

Scale Factor: 1

Value: 25, 40, 55, 70, 60, 20, 5

Duration: 1, 1, 1, 1, 1, 1, 1

Cada par de valores “Value” - “Duration” especifica la frecuencia, medida en entidades por hora, durante el periodo de tiempo indicado. Suponiendo que el día consta de 24 horas (este dato se introduce al definir el experimento), la planificación establece que durante el primer día llegarán aproximadamente $25 \cdot 24 = 600$ entidades, durante el segundo día $40 \cdot 24 = 960$, durante el tercero $55 \cdot 24 = 1320$, y así sucesivamente, de modo que el séptimo día llegarán $5 \cdot 24 = 120$ entidades. □

Generación de procesos de Poisson no estacionarios

Una forma *incorrecta* de modelar un proceso de Poisson no estacionario consiste en definir como tiempo entre llegadas (“Time Between Arrivals”), en el módulo “Create”, una distribución exponencial con un valor para la media igual a una función del tiempo definida por el usuario. Esta forma (incorrecta) de modelar el proceso de Poisson no estacionario proporciona una solución aproximada si la variación de la frecuencia entre intervalos es pequeña. Sin embargo, si la variación es grande, este método proporciona resultados fuertemente erróneos. El siguiente ejemplo pretende ilustrar este problema.

Ejemplo 8.8.2. *Considérese que la variación temporal de la frecuencia se describe empleado únicamente dos intervalos, $k = 2$, de 30 minutos de duración cada uno:*

- *La frecuencia en el primer intervalo es 3 llegadas por hora, es decir, el valor medio del intervalo de tiempo entre llegadas es 20 minutos.*
- *La frecuencia en el segundo intervalo es 60 llegadas por hora, es decir, el tiempo medio entre llegadas es 1 minuto.*

Si la última llegada del primer periodo se produjo en el instante 29 minutos, entonces se generará el siguiente instante de llegada usando el tiempo medio entre llegadas 20 minutos. Usando una distribución exponencial con media 20, es probable obtener un valor mayor que 31 para el siguiente intervalo, con lo cual no se obtendría ninguna llegada durante el segundo intervalo, cuando en realidad el valor esperado del número de llegadas en este segundo intervalo es 30. □

En general, usando este método (erróneo) para generar los procesos de Poisson no estacionarios, se obtiene una disminución incorrecta en el número de llegadas cuando entre un intervalo y el siguiente se produce un aumento en la frecuencia, y se produce un aumento incorrecto en el número de llegadas cuando entre un intervalo y el siguiente se produce una disminución en la frecuencia.

El método empleado por Arena para generar los procesos de Poisson no estacionarios se discutirá en el tema dedicado a la generación de observaciones de variables aleatorias.

8.9. Análisis de las entradas usando Arena

En la práctica, el ajuste de las distribuciones de probabilidad a los datos suele realizarse usando herramientas software¹⁸. En particular, el entorno de simulación Arena soporta las capacidades necesarias para realizar este tipo de análisis, permitiendo que las distribuciones ajustadas (teóricas o empíricas, continuas o discretas) puedan ser usadas directamente en el modelo. A continuación vamos a discutir los fundamentos de la herramienta *Input Analyzer* de Arena, así como su aplicación a algunos ejemplos.

¹⁸Para esta tarea puede usarse software estadístico de propósito general, como por ejemplo Splus o R (la versión GNU de S), o bien paquetes software con este propósito específico. Entre estos últimos, cabe destacar ExpertFit (puede descargarse gratuitamente una versión de demostración de <http://www.averill-law.com/>).

El ajuste de una distribución a los datos experimentales y el empleo de aquella en el modelo se realizan siguiendo los siguientes cuatro pasos:

1. Crear un fichero de texto que contenga las observaciones experimentales.
2. Ajustar una o más distribuciones a los datos.
3. Seleccionar qué distribución desea usarse.
4. Copiar la expresión generada por *Input Analyzer* en el campo apropiado del modelo en Arena.

A continuación se explica de forma detallada cada uno de ellos.

Formato del fichero de datos

El formato del fichero de texto que contiene los datos experimentales x_1, \dots, x_n es completamente libre. Los datos pueden estar separados por un número arbitrario de espacios en blanco, tabuladores, o retornos de carro. Pueden existir tantas observaciones por línea del fichero como se desee y cada línea puede tener un número variable de observaciones. La extensión por defecto del fichero es “.dst”.

Arranque de *Input Analyzer* y carga del fichero de datos

En el menú de Arena, seleccionar:

Tools / Input Analyzer

Con ello se abre la pantalla *Input Analyzer*. A partir de este punto se trabaja en esta pantalla. A continuación hay que crear una nueva sesión de trabajo:

File / New

y cargar el fichero de datos:

File / Data File / Use Existing ...

Los datos cargados pueden visualizarse seleccionando:

Window / Input Data

Histograma

Al cargar el fichero con las muestras experimentales, se muestra automáticamente el histograma de las mismas. El número de clases (intervalos) y el rango del histograma, pueden modificarse seleccionando:

Options / Parameters / Histogram ...

La modificación del rango del histograma permite restringir las muestras que van a ser estudiadas y a las que se va a realizar el ajuste: sólo son consideradas aquellas contenidas en el intervalo indicado. Esta capacidad es útil para el estudio de muestras multimodales.

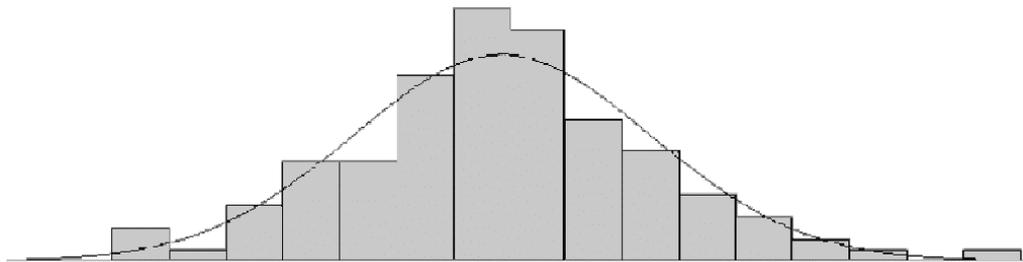


Figura 8.11: Ajuste a los datos de una distribución normal.

Ajuste de una distribución

La opción:

Fit

permite seleccionar qué distribución desea ajustarse de entre determinado conjunto de posibles distribuciones. Como ejemplo, en la Figura 8.11 se muestra el ajuste de una distribución normal a un conjunto de 125 datos experimentales.

La densidad de probabilidad de la distribución ajustada se dibuja superpuesta al histograma. Los datos resultado del ajuste no sólo se muestra en la pantalla, además queda automáticamente grabado en un fichero que lleva el nombre de la familia de la distribución ajustada y tiene extensión “.out”. Esta información es la siguiente:

- *Distribution*. Familia de la distribución ajustada.
- *Expression*. Es la expresión exacta que debe usarse en el modelo de Arena para representar la distribución ajustada. Puede copiarse al modelo mediante: *Edit / Copy Expression*. Esta opción es especialmente útil cuando se ajusta una distribución empírica.
- *Square Error*. Es el error cuadrático medio. Se calcula como el promedio del cuadrado del error cometido en las clases del histograma:

$$\frac{\sum_{i=1}^{k_H} e_i^2}{k_H} \quad (8.61)$$

donde k_H es el número de clases del histograma.

El error cometido en la clase i , e_i , es la diferencia entre la frecuencia relativa observada en esa clase y la frecuencia relativa predicha por la distribución ajustada. Cuanto mayor es el error cuadrático medio, más se desvía la distribución ajustada de los datos observados y, en consiguiente, peor es el ajuste.

- *Chi-square test*. Resultado de la aplicación del test chi-cuadrado para contrastar la bondad del ajuste. Se indica el número de clases del test, el número de grados de libertad¹⁹, el valor del estadístico χ^2 y el correspondiente *p-value*.

El *p-value* es la probabilidad de muestrear de la distribución ajustada un conjunto de observaciones más inconsistente con la distribución que los datos experimentales x_1, \dots, x_n .

¹⁹Obsérvese que Arena compara el estadístico χ^2 con el punto crítico de la distribución χ_{k-m-1}^2 , donde k es el número de clases del test (en general diferente del número de clases del histograma) y m es el número de parámetros de la distribución que se han ajustado usando las observaciones.

Cuanto mayor es el valor del *p-value* (siempre vale entre 0 y 1), mejor es el ajuste. Como en el caso de cualquier test estadístico, obtener un valor elevado del *p-value* no constituye una prueba de la bondad del ajuste, sino una indicación que no existe evidencia para refutar el ajuste.

En el texto (Kelton et al. 2002) puede encontrarse la siguiente recomendación. Si el valor del *p-value* del ajuste de una o varias distribuciones teóricas es suficientemente grande (por ejemplo, mayor que 0.10), entonces se pueden usar estas distribuciones teóricas (siempre que el tamaño de la muestra sea lo suficientemente grande como para considerar significativos los resultados de los tests). Si el valor de los *p-values* de los ajustes de las distribuciones teóricas son bajos, entonces es preferible emplear una distribución empírica.

- *Kolmogorov-Smirnov test*. Resultado obtenido de aplicar el test de Kolmogorov-Smirnov. El analizador de entradas indica el valor del estadístico del test y el correspondiente valor del *p-value*.

Ajuste de todas las distribuciones

Seleccionando:

Fit / Fit All

se realiza automáticamente el ajuste de todas las familias de distribuciones soportadas por *Input Analyzer*. Se genera un informe que contiene el valor del error cuadrático medio obtenido en cada caso. El informe correspondiente a los datos mostrados en la Figura 8.11 es el siguiente:

| Function | Sq Error |
|-------------|----------|
| Beta | 0.00492 |
| Normal | 0.00705 |
| Weibull | 0.0101 |
| Gamma | 0.0155 |
| Erlang | 0.0167 |
| Lognormal | 0.0243 |
| Triangular | 0.0343 |
| Uniform | 0.103 |
| Exponential | 0.138 |

Debe tenerse en cuenta que los valores obtenidos de este error dependen del número de clases del histograma. El informe puede visualizarse seleccionando:

Window / Fit All Summary

Asimismo, los resultados se almacenan en un fichero llamado "*SUMMARY.OUT*". También se muestran los datos del ajuste de la distribución con el menor error cuadrático medio (en el ejemplo anterior, la distribución beta).

En ocasiones existen varias distribuciones, pertenecientes a distintas familias, que se ajustan razonablemente bien a los datos experimentales. En estos casos, pueden considerarse los siguientes factores a la hora de escoger una u otra:

- Si durante el estudio de simulación se planea modificar el valor de los parámetros de esta distribución de entrada, por ejemplo, para realizar estudios de sensibilidad o para evaluar diferentes escenarios, conviene escoger una distribución cuyos parámetros tengan un significado conceptual claro. Por ejemplo, para representar los intervalos

entre llegadas, puede encontrarse que una distribución exponencial y una de Weibull proporcionan ajustes similares a los datos experimentales. En este caso, es mucho más sencillo trabajar (por ejemplo, modificar la frecuencia de llegadas) con la distribución exponencial que con la de Weibull.

- Es una buena práctica evaluar la sensibilidad de los resultados de la simulación frente a variaciones en las distribuciones de entrada. En particular, si existen varias distribuciones candidatas, puede ejecutarse la simulación para cada una de ellas y analizar en qué medida difieren los resultados obtenidos.

EJERCICIOS

8.1 Dibuje en una misma gráfica la densidad de probabilidad de cuatro distribuciones normales con media cero, pero con desviaciones estándar $\frac{1}{4}$, $\frac{1}{2}$, 1 y 2 respectivamente.

8.2 Dibuje la probabilidad de la distribución de Poisson que resulta cuando el parámetro α es igual a:

1. $\alpha = \frac{1}{2}$
2. $\alpha = 1$
3. $\alpha = 2$
4. $\alpha = 4$

8.3 Dibuje en una misma gráfica la densidad de probabilidad de dos distribuciones exponenciales, la primera con $\lambda = 0.6$, y la segunda con $\lambda = 1.2$.

8.4 En una línea de ensamblaje de coches, se emplea un robot para colocar las puertas. Se sospecha que el tiempo de instalación obedece una distribución normal. El robot es capaz de medir de manera precisa el tiempo que tarda en instalar cada puerta. Se tomaron automáticamente 20 medidas del tiempo de instalación, obteniéndose los siguientes tiempos, medidos en segundos:

| | | | |
|--------|--------|--------|--------|
| 99.79 | 99.56 | 100.17 | 100.33 |
| 100.26 | 100.41 | 99.98 | 99.83 |
| 100.23 | 100.27 | 100.02 | 100.47 |
| 99.55 | 99.62 | 99.65 | 99.82 |
| 99.96 | 99.90 | 100.06 | 99.85 |

Dibuje el histograma de los datos experimentales y realice el ajuste de una distribución normal usando la herramienta *Input Analyzer*.

Dibuje el gráfica Q-Q comparando los datos experimentales con una distribución normal. ¿Soporta el gráfico Q-Q la hipótesis de que los datos están distribuidos normalmente?

8.5 Se ha contabilizado el número de vehículos que llegan a determinado semáforo durante un periodo de 5 minutos, entre las 7:00 AM y las 7:05 AM. Las medidas se han realizado todos los días laborables, durante 20 semanas. Las observaciones obtenidas son las siguientes:

| Llegadas por periodo | Frecuencia | Llegadas por periodo | Frecuencia |
|-------------------------|------------|-------------------------|------------|
| 0 | 12 | 6 | 7 |
| 1 | 10 | 7 | 5 |
| 2 | 19 | 8 | 5 |
| 3 | 17 | 9 | 3 |
| 4 | 10 | 10 | 3 |
| 5 | 8 | 11 | 1 |

Dibujar el histograma de los datos experimentales y realizar el ajuste a una distribución de Poisson. A continuación, aplicad el test chi-cuadrado para contrastar la hipótesis de que las observaciones experimentales obedecen la distribución de Poisson ajustada.

8.6 En un proceso de llegada de entidades a un sistema, se han registrado los intervalos de tiempo entre llegadas durante un periodo de 100 minutos. Las observaciones registradas han sido las siguientes (ordenadas en el orden en que se han recogido):

| | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 0.44 | 0.53 | 2.04 | 2.74 | 2.00 | 0.30 | 2.54 | 0.52 | 2.02 | 1.89 | 1.53 | 0.21 |
| 2.80 | 0.04 | 1.35 | 8.32 | 2.34 | 1.95 | 0.10 | 1.42 | 0.46 | 0.07 | 1.09 | 0.76 |
| 5.55 | 3.93 | 1.07 | 2.26 | 2.88 | 0.67 | 1.12 | 0.26 | 4.57 | 5.37 | 0.12 | 3.19 |
| 1.63 | 1.46 | 1.08 | 2.06 | 0.85 | 0.83 | 2.44 | 2.11 | 3.15 | 2.90 | 6.58 | 0.64 |

Contraste la hipótesis de que los intervalos entre llegadas están distribuidos exponencialmente, empleando para ello el test de Kolmogorov-Smirnov.

- 8.7** Entre los ejercicios del Capítulo 6, se propuso el modelado de un servicio de atención telefónica completamente automatizado, cuya función es informar al cliente acerca de tres tipos de producto: A, B y C. La llegada de llamadas al sistema era un proceso de Poisson estacionario, con una frecuencia de una llamada por minuto.

Considérese ahora que el sistema funciona durante 10 horas al día (de 8 AM a 6 PM), y que la *frecuencia de llegada de llamadas* al sistema varía a lo largo del día (lo cual es muy típico en este tipo de sistemas). En la tabla siguiente se muestra el valor de la frecuencia, expresada en llamadas por hora, para cada periodo de 30 minutos durante el cual el servicio está en funcionamiento.

| Periodo | Frec. | Periodo | Frec. | Periodo | Frec. | Periodo | Frec. |
|---------------|-------|---------------|-------|-------------|-------|-------------|-------|
| 8:00 - 8:30 | 20 | 10:30 - 11:00 | 75 | 1:00 - 1:30 | 110 | 3:30 - 4:00 | 90 |
| 8:30 - 9:00 | 35 | 11:00 - 11:30 | 75 | 1:30 - 2:00 | 95 | 4:00 - 4:30 | 70 |
| 9:00 - 9:30 | 45 | 11:30 - 12:00 | 90 | 2:00 - 2:30 | 105 | 4:30 - 5:00 | 65 |
| 9:30 - 10:00 | 50 | 12:00 - 12:30 | 95 | 2:30 - 3:00 | 90 | 5:00 - 5:30 | 45 |
| 10:00 - 10:30 | 70 | 12:30 - 1:00 | 105 | 3:00 - 3:30 | 85 | 5:30 - 6:00 | 30 |

Condiciones iniciales: al iniciarse el día, no hay ninguna llamada en el sistema.

Condición de finalización: desea simularse el sistema durante las 10 horas que permanece en funcionamiento cada día.

Objetivo del estudio: estimar el número total de "llamadas perdidas", es decir, el número total de llamadas que no han podido ser atendidas por encontrarse todas las líneas ocupadas.

- 8.8** Usando la herramienta *Input Analyzer*, abra una nueva ventana y genere un nuevo fichero de datos que contenga 50 observaciones de una distribución Erlang con parámetros: $\text{ExpMean} = 12$, $k = 3$ y $\text{Offset} = 5$. Para ello debe usarse la opción: *File / Data File / Generate New*.

Una vez obtenido el fichero, ejecute: *Fit / Fit All*, para obtener el mejor ajuste entre las distribuciones disponibles. Repita el proceso para 500, 5000 y 25000 observaciones, usando los mismos parámetros de la distribución Erlang. Compare los resultados del ajuste para las cuatro muestras de datos.

Tema 9

Generación de números aleatorios

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los conceptos fundamentales de la generación de números aleatorios y del contraste de los generadores.
- Comparar métodos para la generación de números aleatorios.
- Contrastar métodos para la generación de números aleatorios, mediante la aplicación de tests empíricos.

9.1. Introducción

La generación de números aleatorios es un tema complejo y de suma importancia, ya que constituye el corazón de la simulación estocástica. La simulación de cualquier modelo que contiene componentes aleatorios requiere de la aplicación de un método para la generación de muestras independientes de variables aleatorias, a partir de determinadas distribuciones de probabilidad¹. Estas muestras son obtenidas mediante determinadas transformaciones (dependientes de la distribución de probabilidad) a partir de números distribuidos $U(0, 1)$. Estos números independientes, distribuidos $U(0, 1)$ reciben el nombre de *números aleatorios*.

Ejemplo 9.1.1. *Para ilustrar el empleo de los números aleatorios en la generación de observaciones de variables aleatorias, consideremos el lanzamiento de un dado ideal con las caras numeradas entre 1 y 6. Si se dispone de una fuente de números aleatorios, podría simularse el lanzamiento del dado asociando los números aleatorios con las caras del dado de la forma indicada en la Figura 9.1. Análogamente, podría simularse el lanzamiento de un dado “trucado”, simplemente cambiando la probabilidad asociada a cada una de las caras del mismo.* □

¹En los textos escritos en lengua inglesa se distingue entre “*random variables*” y “*random variates*”. Una variable aleatoria (*random variable*) se define en la teoría de probabilidad como una función que satisface determinadas propiedades (ver Apéndice A), con lo cual no es riguroso hablar de “generar variables aleatorias”. De forma más precisa, debe hablarse de “generar observaciones de variables aleatorias”. A las observaciones (o realizaciones) de las variables aleatorias es a lo que se denomina “*random variates*”.

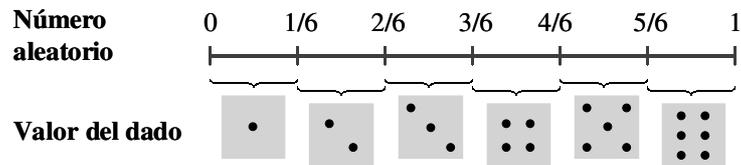


Figura 9.1: Simulación del lanzamiento de un dado.

Existen dos grandes clases de generadores de números aleatorios: los *generadores físicos* y los *generadores aritméticos*. La elección acerca de cuál de los dos resulta más adecuado depende de la aplicación.

- *Generadores físicos*. Son dispositivos físicos que emplean fuentes externas, como son la desintegración de un material radioactivo o el ruido eléctrico, para generar números aleatorios. Los más frecuentemente usados están basados en circuitos eléctricos dotados de una fuente de ruido (frecuentemente una resistencia o un diodo semiconductor) que es amplificada, muestreada y comparada con una señal de referencia para producir secuencias de bits. Por ejemplo, si el ruido es menor que la referencia se produce un 0 y si es mayor un 1. Frecuentemente, en las secuencias de bits así obtenidas los 1's no tienen la misma probabilidad que los 0's, con lo cual deben ser post-procesadas para obtener secuencias de bits aleatorios. Para su aplicación, estos bits aleatorios se unen para formar bytes, números enteros o números reales, según sea preciso.
- *Generadores aritméticos*. Son algoritmos deterministas que son ejecutados mediante el ordenador. Los buenos generadores aritméticos producen secuencias de números que son indistinguibles de realizaciones independientes de variables uniformes. Existen dos grandes familias de generadores aritméticos: los generadores *lineales* y los *no lineales*. En simulación los generadores más comúnmente usados son los lineales.

Se han desarrollado *tests estadísticos* para contrastar la bondad de los generadores, es decir, para contrastar si la secuencia de números generada puede “confundirse” con observaciones independientes de la distribución $U(0, 1)$.

Los entornos de simulación, y en general cualquier software estadístico, soportan uno o varios tipos de generadores de números aleatorios. Es muy recomendable identificar de qué tipo de generador se trata y cuáles son sus parámetros, con el fin de, a no ser que sea un generador de reconocidas buenas propiedades, someterlo a una batería de tests estadísticos.

Existen dos tipos diferentes de tests: los *tests empíricos* y los *tests teóricos*.

- *Tests empíricos*. Permiten analizar una secuencia de números obtenida del generador, con el fin de calcular con qué probabilidad cabría haberla obtenido muestreando una distribución $U(0, 1)$. El resultado del test sólo tiene validez para la secuencia analizada, no pudiéndose afirmar nada acerca de otras secuencias, que pudieran ser obtenidas del mismo generador, sobre las que no se ha aplicado el test.
- *Tests teóricos*. Analizan el generador basándose en el valor de sus parámetros, sin necesidad de generar un solo número aleatorio. Sus resultados son globales, es decir, son aplicables a cualquier secuencia obtenida del generador. Se trata de tests sofisticados y matemáticamente complejos.

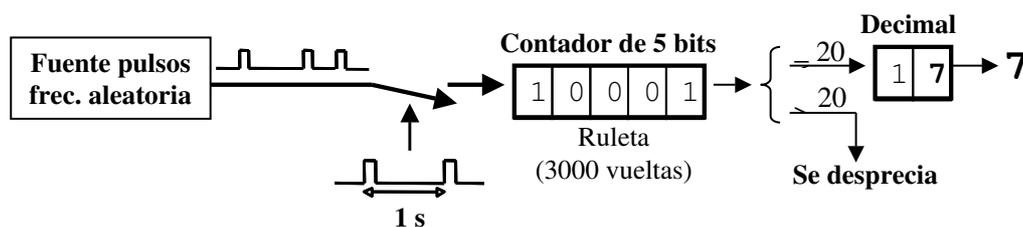


Figura 9.2: Generación de un dígito aleatorio mediante el método de RAND Corporation.

9.2. Generadores físicos

La metodología para la generación de números aleatorios tiene una historia larga e interesante². Los métodos más antiguos se realizaban esencialmente a mano: tirar dados, usar cartas, extraer bolas de una urna “perfectamente agitada”, etc.

Durante la primera mitad del siglo XX, se construyeron dispositivos mecánicos y eléctricos para la generación de números aleatorios, tales como discos giratorios o circuitos eléctricos basados en tubos de vacío con pulsaciones aleatorias. Durante la segunda mitad del siglo XX aparecieron gran cantidad de trabajos proponiendo generadores físicos de números aleatorios. En algunos casos, los números aleatorios eran publicados en forma de tablas, de modo que pudieran ser usados en la aplicación de los métodos de Monte Carlo por otros investigadores.

La tabla de números aleatorios más ampliamente usada ha sido la tabla generada por *RAND Corporation*³ en el año 1955. Una de las ventajas de esta tabla frente a las disponibles hasta aquella fecha era su gran tamaño: un millón de dígitos aleatorios.

Disponer de secuencias suficientemente largas de números aleatorios resultaba esencial. La razón era que, cuando el tamaño de la tabla era insuficiente, ésta debía reutilizarse varias veces durante la simulación, con el consiguiente peligro de introducir correlaciones que falsearan los resultados del estudio.

La tabla de 1 millón de dígitos aleatorios de RAND Corporation

Los dígitos aleatorios de la tabla se obtuvieron aleatorizando los dígitos obtenidos de un circuito eléctrico cuyo funcionamiento se asemejaba al de una ruleta⁴. El circuito consistía en una fuente de pulsos de frecuencia aleatoria, que proporcionaba un promedio de 100.000 pulsos por segundo.

Una vez por segundo se conectaba este tren de pulsos aleatoriamente espaciados a un contador de 5 bits. La cuenta se realizaba durante un intervalo de tiempo constante, inferior a 1 segundo. Una vez el contador había alcanzado su valor máximo (es decir, 31), al recibir el siguiente pulso adquiría el valor 0, a continuación el 1 y así sucesivamente.

Esta cuenta cíclica se asemeja al comportamiento de una ruleta de 32 números que gira repetidas veces, recorriendo todos los números, hasta que finalmente se detiene en uno de ellos.

²En (Law & Kelton 2000) se citan referencias bibliográficas acerca de este tema.

³RAND Corporation (1955): A Million Random Digits with 100,000 Normal Deviates. Free Press.

La tabla puede descargarse gratuitamente de la dirección:

<http://www.rand.org/publications/MR/MR1418/>

⁴Puede encontrarse una descripción detallada en:

<http://www.rand.org/publications/classics/randomdigits/>

TABLE OF RANDOM DIGITS 1

| | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 00000 | 10097 | 32533 | 76520 | 13586 | 34673 | 54876 | 80959 | 09117 | 39292 | 74945 |
| 00001 | 37542 | 04805 | 64894 | 74296 | 24805 | 24037 | 20636 | 10402 | 00822 | 91665 |
| 00002 | 08422 | 68953 | 19645 | 09303 | 23209 | 02560 | 15953 | 34764 | 35080 | 33606 |
| 00003 | 99019 | 02529 | 09376 | 70715 | 38311 | 31165 | 88676 | 74397 | 04436 | 27659 |
| 00004 | 12807 | 99970 | 80157 | 36147 | 64032 | 36653 | 98951 | 16877 | 12171 | 76833 |
| 00005 | 66065 | 74717 | 34072 | 76850 | 36697 | 36170 | 65813 | 39885 | 11199 | 29170 |
| 00006 | 31060 | 10805 | 45571 | 82406 | 35303 | 42614 | 86799 | 07439 | 23403 | 09732 |
| 00007 | 85269 | 77602 | 02051 | 65692 | 68665 | 74818 | 73053 | 85247 | 18623 | 88579 |
| 00008 | 63573 | 32135 | 05325 | 47048 | 90553 | 57548 | 28468 | 28709 | 83491 | 25624 |
| 00009 | 73796 | 45753 | 03529 | 64778 | 35808 | 34282 | 60935 | 20344 | 35273 | 88435 |
| 00010 | 98520 | 17767 | 14905 | 68607 | 22109 | 40558 | 60970 | 93433 | 50500 | 73998 |
| 00011 | 11805 | 05431 | 39808 | 27732 | 50725 | 68248 | 29405 | 24201 | 52775 | 67851 |
| 00012 | 83452 | 99634 | 06288 | 98083 | 13746 | 70078 | 18475 | 40610 | 68711 | 77817 |
| 00013 | 88685 | 40200 | 86507 | 58401 | 36766 | 67951 | 90364 | 76493 | 29609 | 11062 |

Figura 9.3: Comienzo de la tabla de un millón de dígitos aleatorios de RAND Corporation.

El tiempo durante el cual se contaban los pulsos fue calculado de modo que en promedio el contador daba aproximadamente 3000 vueltas. Una vez finalizado el intervalo de tiempo de cuenta de los pulsos, el número almacenado en el contador (comprendido entre 0 y 31) era convertido a decimal. Cuando el número era mayor que 20 se descartaba. Si era menor que 20, el dígito menos significativo del número constituía el dígito aleatorio. Este dígito era almacenado grabándolo en una tarjeta perforada. En la Figura 9.2 se representa esquemáticamente el proceso de generación de un dígito aleatorio.

Los circuitos tuvieron que ser modificados y ajustados repetidamente, hasta que la secuencia de dígitos obtenida adquirió unas propiedades estadísticas adecuadas, es decir, era capaz de superar determinados tests estadísticos. En la Figura 9.3 se muestra el principio de la tabla de RAND Corporation. Las líneas de la tabla están numeradas desde 00000 hasta 19999. Para usar la tabla, hay que determinar una posición de comienzo al azar y leer la tabla de la misma forma que se leería un libro. Puede aumentarse el tamaño de la tabla modificando la dirección de lectura al reutilizarla en una misma simulación: leyendo por columnas en lugar de por líneas, leyendo la tabla hacia atrás, leyendo las hojas de arriba abajo y las líneas de derecha a izquierda, etc.

Generador de bits aleatorios de Vincent

Otro generador físico de gran importancia histórica es el generador de bits aleatorios de Vincent. En el año 1970, C. Vincent propuso el siguiente método para la generación de bits aleatorios⁵: contar el número de pulsos generados aleatoriamente durante un determinado intervalo de tiempo y, en función de que la suma obtenida sea par o impar, seleccionar 0 ó 1 respectivamente. Los pulsos pueden obtenerse midiendo con un detector una fuente radioactiva o bien, más convenientemente, mediante un circuito que detecte los picos que superan determinado valor, conectado a la salida de una fuente de ruido.

⁵Para profundizar en el tema, pueden consultarse las referencias siguientes:

Vincent, C. (1970): "The generation of truly random binary numbers". *Journal of Physics E*. 3(8): 594-598.

Vincent, C. (1971): "Precautions for accuracy in the generation of truly random binary numbers". *Journal of Physics E*. 4(11): 825-828.

Vincent, C. (1973): "Random pulse trains, their measurement and statistical properties". *Institution of Electrical Engineers (I.E.E.) monograph series*.

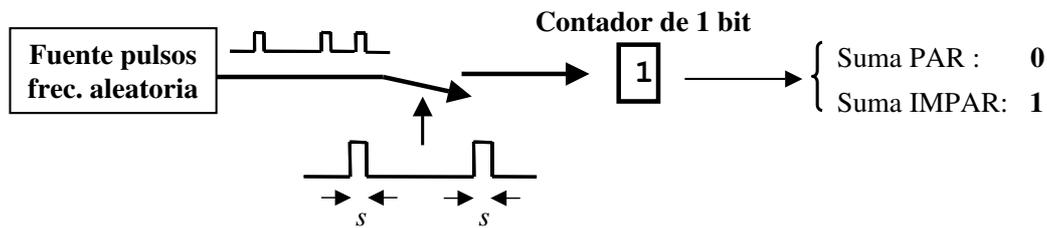


Figura 9.4: Esquema del funcionamiento del generador de Vincent.

En la figura 9.4 se muestra un esquema del procedimiento para la generación de bits aleatorios. La máquina generadora de números aleatorios de Vincent satisface que el número de pulsos k , contados en un intervalo de tiempo s , está distribuido de acuerdo a una distribución de Poisson con media $\lambda \cdot s$:

$$\text{Prob} \{ (N(t+s) - N(t)) = k \} = \frac{e^{-\lambda \cdot s} \cdot (\lambda \cdot s)^k}{k!} \quad (9.1)$$

La diferencia entre la probabilidad de que el número de pulsos sea par, y la probabilidad de que el número de pulsos sea impar, es igual a:

$$\text{Prob} \{ k \text{ es par} \} - \text{Prob} \{ k \text{ es impar} \} = \sum_{k:0}^{\infty} (-1)^k \cdot \frac{e^{-\lambda \cdot s} \cdot (\lambda \cdot s)^k}{k!} = e^{-2 \cdot \lambda \cdot s} \quad (9.2)$$

Dado que $e^{-2 \cdot \lambda \cdot s}$ vale aproximadamente 10^{-13} para un valor de $\lambda \cdot s$ tan pequeño como 15, la probabilidad de obtener un 0 es prácticamente la misma que la de obtener un 1, con lo cual el método de Vincent es muy satisfactorio.

Cualquier deriva que ocurriera en el detector de picos, o en el nivel de ruido de la fuente, únicamente afectaría al valor de $\lambda \cdot s$ (es decir, a la media de pulsos recibidos en el intervalo de longitud s), con lo cual tendría un efecto extremadamente pequeño en las probabilidades de obtener 0's y 1's.

Secuencias de pulsos aleatoriamente espaciados

Diodo Zener. En 1978, Yarza y Martínez⁶ propusieron un circuito para generar pulsos aleatoriamente espaciados, usando un *diodo Zener*, polarizado en el codo de su curva característica, como fuente de ruido. La señal de ruido era amplificada y comparada con un nivel umbral que actuaba como señal de control. Variando el valor de esta señal de control, puede modificarse la frecuencia promedio de los pulsos. Finalmente, la señal aleatoria binaria que se obtiene a la salida del comparador es muestreada por el tren de pulsos del reloj, obteniéndose un tren de pulsos aleatorio. En teoría, el diodo polarizado en inversa debería ser capaz de aleatoriedad perfecta, dado que su respuesta está determinada por múltiples eventos aleatorios a escala atómica.

Descomposición de muestras radioactivas. Existen procedimientos, basados en la medida de la emisión de muestras radioactivas, para generar un tren de pulsos con la siguiente propiedad: el intervalo de tiempo entre cada dos pulsos consecutivos es aleatorio, e independiente de los intervalos de tiempo entre los demás pulsos.

⁶Yarza, A and Martínez, P. (1978): "A true random pulse train generator". Electronic Engineering. 50(614): 21-23

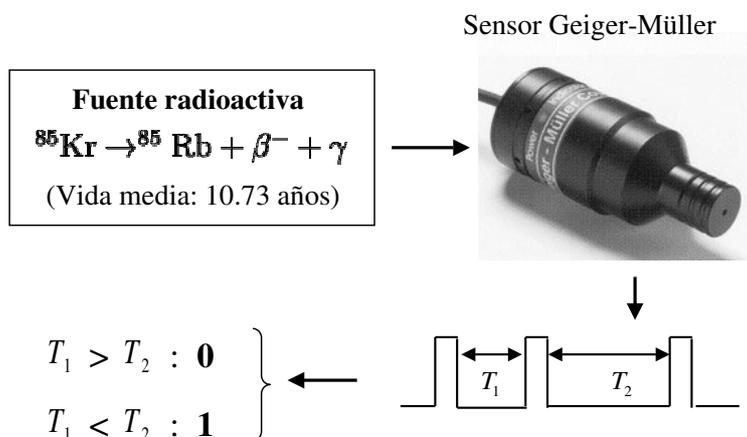


Figura 9.5: Funcionamiento del generador del proyecto *Hotbit*.

Un generador de este tipo se consigue midiendo la transición de Kriptón-85 a Rubidio. Esta transición se produce espontáneamente, de tal modo que la vida media del Kriptón-85 son 10.73 años. Esto significa que si se tiene una muestra con 100 millones de átomos de Kriptón-85, transcurridos 10.73 años la mitad de ellos se habrán transformado en átomos de Rubidio. Pese a este hecho, no existe forma de predecir cuando un determinado átomo a Kriptón-85 sufrirá la transformación, y en eso se basa el procedimiento para la generación de números aleatorios, ya que dada una muestra de átomos de Kriptón-85, no hay forma de predecir en qué preciso instante se producirá la siguiente transición.

El instante en que se produce la emisión radioactiva puede detectarse mediante un tubo Geiger-Müller conectado a un ordenador. El tubo Geiger-Müller emite un pulso cada vez que detecta una emisión. En la figura 9.5 se muestra un esquema del método.

Un procedimiento para generar bits aleatorios a partir del tren de pulsos (de frecuencia variable) obtenido es el propuesto en el proyecto *HotBit*⁷. Para generar un bit se espera hasta que ocurra la primera cuenta y se mide el tiempo, T_1 , que transcurre hasta que se produce la siguiente. Se espera hasta que se produzca un tercer pulso, y se mide el tiempo transcurrido, T_2 , desde el segundo pulso.

De este modo se obtiene el valor de dos intervalos de tiempo aleatorios: T_1 y T_2 . Si ambos intervalos son iguales, se desprecia la medida y se intenta de nuevo. Si T_1 es menor que T_2 , entonces el bit generado es un 0, en caso contrario, un 1. En la práctica, con el fin de evitar algún posible sesgo en los aparatos de medida, se intercambia el signo de la comparación con cada bit generado.

Señales de radio. Otro procedimiento para obtener secuencias de pulsos aleatoriamente espaciados entre sí es sintonizando un receptor de radio a una frecuencia en la que nadie está emitiendo. Un ejemplo de ello es el siguiente procedimiento, empleado en el proyecto *Random.org*⁸. La señal del ruido atmosférico recibido por el receptor es introducida en un ordenador, donde es muestreada y digitalizada a una palabra de 8 bits. Los 7 bits más significativos de cada palabra son descartados, mientras que el bit menos significativo es almacenado.

La secuencia de bits obtenida de esta forma debe ser “corregida”, a fin de garantizar que el número aproximado de 0's y 1's sea el mismo. Existe varios *algoritmos para la corrección del sesgo*. Cabe destacar uno de ellos, que es debido a von Neumann, y elimina completamente

⁷Fourmilab, Suiza. <http://www.fourmilab.ch/hotbits/>

⁸<http://www.random.org/>

el sesgo hacia 0 o hacia 1 de los datos. Consiste en los pasos siguientes. Se leen los bits de dos en dos.

- Si existe una transición entre sus valores (es decir, los bits son 01 ó 10), el primero de ellos se considera un bit aleatorio y el segundo se descarta.
- Si no existe transición entre los dos bits (es decir, son 00 ó 11), entonces se descartan los dos bits y se leen los dos siguientes.

Aplicación a la simulación

El interés en la construcción de “máquinas” generadoras de números aleatorios sigue ciertamente vigente hoy en día. No obstante, la aplicación de los generadores físicos en los estudios de simulación presenta algunas desventajas:

- Podría conectarse directamente la máquina generadora de números aleatorios al ordenador. La desventaja de esta técnica es que no permite reproducir exactamente las secuencias de números aleatorios previamente generadas. Como se verá, esto es necesario para poder aplicar las técnicas de reducción de la varianza.
- Podrían almacenarse en una tabla los números generados por este tipo de máquinas y leer la tabla durante la simulación. Sin embargo, para poder aplicar esta técnica a estudios de simulación complejos, en los que se requiere millones de números aleatorios, hace falta usar memorias de gran capacidad. Por otra parte, la lectura de los datos desde la tabla hace que la ejecución de la simulación sea más lenta.

9.3. Generadores aritméticos

La ventaja del empleo de tablas de números aleatorios es su reproducibilidad. La desventaja es la lentitud y el riesgo de agotar la tabla.

En vista de estas dificultades, John von Neumann sugirió en el año 1951 un procedimiento aritmético para la generación de números aleatorios: el *método de los cuadrados medios*, que puede considerarse el precursor de los generadores aritméticos actuales.

El método para obtener un nuevo número es extraer los dígitos centrales del resultado de elevar al cuadrado el último número de la secuencia. Por ejemplo, si se están generando números de 4 dígitos, el número siguiente a 5232 se obtiene elevando éste al cuadrado, $5232^2 = 27373824$, y extrayendo los 4 dígitos centrales, es decir, 3738.

Análisis posteriores han demostrado que el método de los cuadrados medios no tiene buenas propiedades estadísticas, con lo que no debe ser empleado.

Una característica de los métodos aritméticos es que son *secuenciales*: cada nuevo número se determina a partir de uno o varios de sus predecesores aplicando una determinada fórmula matemática.

Los números generados mediante los *generadores aritméticos* no son aleatorios, si se entiende como tal que sean impredecibles. Conocido el algoritmo empleado para generar los números y el estado interno del mismo, se pueden predecir todos los números obtenidos en las siguientes llamadas al algoritmo. Por el contrario, con números genuinamente aleatorios, el conocimiento de un número, o de una secuencia arbitrariamente larga de números, resulta inútil para predecir el valor del siguiente número generado. Por este motivo, los números “aleatorios” generados mediante el ordenador usando generadores aritméticos se denominan “*números pseudo aleatorios*”.

Algunos generadores aritméticos tienen una gran calidad, entendiendo como tal que las secuencias de números obtenidas “parecen” muestras independientes de una distribución uniforme, $U(0,1)$. Es decir, superan todos los tests estadísticos ideados para distinguir secuencias aleatorias de aquellas que contienen algún patrón u orden interno. La hipótesis nula de estos tests es la siguiente:

H_0 : los números son muestras independientes de una distribución $U(0,1)$

La dificultad fundamental de la generación de números aleatorios, ya sea mediante técnicas físicas o aritméticas, es que *mientras existen muchas maneras de probar que una determinada secuencia de números no es aleatoria, no existe ninguna forma de probar que sí lo es.*

Dos conceptos fundamentales, en la generación de números aleatorios, son los de *uniformidad e independencia*.

Uniformidad

Para decidir acerca de la *uniformidad* de una secuencia de números, u_1, u_2, \dots , se puede seguir la siguiente aproximación cualitativa⁹:

- Los números, cuando se representan como puntos en un segmento de línea que va del 0 al 1, deben estar aproximadamente distribuidos uniformemente.
- Las sucesivas parejas $\{u_1, u_2\}, \{u_3, u_4\}, \dots$, cuando se tratan como puntos en el cuadrado de lado unidad, deben estar aproximadamente uniformemente distribuidos.
- ...
- Las sucesivas tuplas de n números $\{u_1, u_2, \dots, u_n\}, \{u_{n+1}, u_{n+2}, \dots, u_{2n}\}, \dots$, cuando se tratan como puntos en un hipercubo n -dimensional, deben estar aproximadamente distribuidos uniformemente.

Una generador puede considerarse suficientemente uniforme a efectos prácticos si su salida satisface estas n condiciones, para un valor de n suficientemente grande para la aplicación en cuestión. En la práctica, la carga computacional de aplicar tests empíricos generalmente es prohibitiva para valores de n por encima de 10.

Independencia

Una posible definición de *independencia* es la siguiente¹⁰: una serie es independiente si no puede ser especificada mediante un algoritmo que requiera menos bits que la serie en sí. Por ejemplo, la serie de bits:

01010000001010011011...

es independiente si la forma más rápida de describirla es simplemente escribirla. Por el contrario, la serie de bits:

0101010101010101...

puede ser descrita mediante un algoritmo más corto que la serie en sí. Esta definición de independencia, aunque es interesante desde el punto de vista matemático, no sugiere ningún método para el diseño de generadores de números aleatorios.

Características deseables de un generador

Desde un punto de vista práctico, *con vistas a su aplicación a la simulación*, existen propiedades de los generadores más importantes que la "aleatoriedad". Algunas de éstas son:

- *Propiedad 1.* Los números generados deben "parecer" estar distribuidos uniformemente en el intervalo $[0, 1]$, y no debe mostrar ninguna correlación entre ellos. Si no se cumplen estas dos condiciones (uniformidad e independencia), los resultados de la simulación pueden ser completamente erróneos.

⁹Descrita en el texto (Bratley et al. 1987).

¹⁰Extraída del texto (Bratley et al. 1987).

- *Propiedad 2.* Desde un punto de vista práctico, el generador debe ser rápido y no precisar de grandes capacidades de almacenamiento de datos.
- *Propiedad 3.* Debe ser posible reproducir exactamente la secuencia de números aleatorios, al menos por dos razones. La primera es que en ocasiones esto facilita el depurado de los programas. La segunda, y más importante, es que frecuentemente se requiere el uso de exactamente la misma secuencia de números aleatorios para simular diferentes modelos, a fin de obtener una comparación más precisa entre ellos.
- *Propiedad 4.* El generador debe permitir la generación de varias secuencias separadas de números aleatorios. Una *secuencia* es un segmento de los números producidos por el generador: una secuencia termina donde comienza la siguiente. De este modo, puede dedicarse una secuencia diferente para cada fuente de aleatoriedad del modelo. Por ejemplo, una secuencia para los intervalos entre llegadas de un tipo de entidad, otra secuencia para los tiempos de proceso de un determinado recurso, etc.

Como se observa, la *reproducibilidad* de la secuencia de números es una característica deseable cuando se usa en simulación, y en este sentido los números pseudo aleatorios son adecuados para este propósito. No es éste el caso, por ejemplo, de las aplicaciones en criptografía, en las cuales es fundamental que la secuencia de números sea impredecible.

La mayoría de los generadores aritméticos son muy rápidos, requieren de muy poca capacidad de almacenamiento de datos y puede reproducirse fácilmente una determinada secuencia de números pseudo aleatorios. De este modo, las Propiedades 2 y 3 son casi universalmente satisfechas. Más aun, muchos generadores (entre los que se encuentran los incluidos en los entornos de simulación) soportan la capacidad de producir múltiples secuencias, con lo que satisfacen la Propiedad 4. Sin embargo, hay muchos generadores que no satisfacen la Propiedad 1 (uniformidad e independencia), que es absolutamente necesaria para obtener resultados correctos de la simulación.

A continuación se van a describir algunos de los generadores aritméticos más comúnmente usados, y posteriormente se discutirán algunos tests empíricos de uniformidad e independencia.

9.4. Generadores congruenciales lineales

Los *generadores congruenciales lineales* fueron introducidos por D. H. Lehmer, en el año 1951. Hoy en día es el tipo de generador aritmético más frecuentemente usado en simulación. La secuencia de números pseudo aleatorios se genera de la forma siguiente:

1. Se obtiene una secuencia de números enteros aplicando la fórmula:

$$z_i = (a \cdot z_{i-1} + c) \pmod{m} \text{ para } i : 1, 2, \dots \quad (9.3)$$

El valor de los parámetros m (módulo), a (multiplicador) y c (incremento) diferencia un generador de otro, determinando también la calidad estadística del generador. Estos parámetros deben ser números enteros, no negativos.

Debe satisfacerse que el valor del módulo (m), sea mayor que el multiplicador ($m > a$), que el incremento ($m > c$) y que el valor inicial con el que se inicia la secuencia ($m > z_0$). Este valor inicial se denomina "semilla".

2. Puesto que z_i se define como el resto de la división de $a \cdot z_{i-1} + c$ por m , se satisface: $0 \leq z_i \leq m - 1$ para $i : 1, 2, \dots$

Para obtener números en el intervalo $[0, 1)$, es preciso dividir los enteros obtenidos por m :

$$u_i = \frac{z_i}{m} \text{ para } i : 1, 2. \quad (9.4)$$

La lógica que subyace en este algoritmo es que, en la multiplicación y en la división, los dígitos de orden más bajo son más difíciles de predecir, no exhibiendo un patrón o una secuencia evidente.

En función del valor del incremento, los generadores congruenciales lineales (abreviado: GCL) se clasifican en los siguientes dos tipos:

- *Mixtos*: aquellos con el incremento diferente de cero: $c > 0$.
- *Multiplicativos*: aquellos con el incremento igual a cero: $c = 0$.

Pueden plantearse las dos siguientes objeciones a los generadores congruenciales lineales:

1. La primera es común a todos los generadores aritméticos: los números generados mediante la Ecuación (9.3) no son aleatorios, ya que son predecibles. En efecto, cualquier número z_i de la secuencia queda completamente fijado, una vez se han establecido los valores de m , a , c y de la semilla (z_0):

$$z_i = \left[a^i \cdot z_0 + \frac{c \cdot (a^i - 1)}{a - 1} \right] \pmod{m} \quad (9.5)$$

Como se ha discutido anteriormente, en realidad esto no es una desventaja en las aplicaciones en simulación, sino una ventaja.

2. La segunda objeción es que los números pseudo aleatorios, u_i , sólo pueden tomar valores entre los números racionales siguientes:

$$\left\{ 0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m} \right\} \quad (9.6)$$

En la práctica esto no supone una limitación seria, ya que el valor de m puede escogerse muy grande, de modo de que el conjunto de valores posibles sea muy denso en el intervalo $[0, 1)$.

Periodo del generador

El concepto de *periodo de un generador* se explica mediante el siguiente ejemplo.

Ejemplo 9.4.1. *Considérese el siguiente generador congruencial lineal¹¹:*

$$z_i = (5 \cdot z_{i-1} + 3) \pmod{16} \quad (9.7)$$

Se escoge el siguiente valor de la semilla $z_0 = 7$. La secuencia formada por los primeros 33 números generados es la siguiente:

```

7 6 1 8 11 10 5 12 15 14 9 0 3 2 13 4
7 6 1 8 11 10 5 12 15 14 9 0 3 2 13 4
7 6
```

Obsérvese que:

$$z_1 = z_{17} = z_{33} = 6$$

$$z_2 = z_{18} = 1$$

... y así sucesivamente.

□

¹¹Se trata únicamente de un ejemplo académico. En la práctica nadie usaría un GCL con un valor de m tan pequeño.

Este comportamiento cíclico es inevitable: si se genera un valor que ya ha aparecido previamente, entonces se obtiene la misma secuencia de valores que se obtuvo previamente, y el ciclo se repite sin fin. Esto es debido a que z_i sólo depende del valor anterior en la secuencia, z_{i-1} .

La longitud del ciclo se llama *periodo del generador*, y se representa: p .

Puesto que $0 \leq z_i \leq m - 1$, se satisface: $p \leq m$.

Generadores de periodo completo

Cuando $p = m$, se dice que el generador congruencial lineal tiene *periodo completo*. Éste es el caso del generador del ejemplo anterior.

En un generador con periodo completo, cualquier semilla z_0 que se escoja de entre los números enteros $\{0, 1, \dots, m - 1\}$, produce el ciclo entero en algún orden determinado. En cambio, en los generadores que no tienen periodo completo, la longitud del ciclo depende de la elección de la semilla.

Es deseable que los generadores tengan periodo completo. En particular, el hecho de que cada entero entre 0 y $m - 1$ aparezca exactamente una vez en cada ciclo, contribuye a la uniformidad de los números generados. Pese a esto, los generadores con periodo completo pueden presentar comportamiento no uniforme en segmentos dentro de ciclo. Por ejemplo, una secuencia de números de longitud $\frac{m}{2}$ puede no estar uniformemente distribuida, dejando grandes “huecos” en el conjunto de valores posibles $\{0, 1, \dots, m - 1\}$.

A continuación se enuncia un teorema que indica las condiciones necesarias y suficientes que debe satisfacer cualquier generador congruencial lineal para tener periodo completo.

Teorema 9.4.1. *Un generador congruencial lineal, definido mediante la fórmula:*

$$z_i = (a \cdot z_{i-1} + c) \pmod{m} \quad (9.8)$$

tiene periodo completo si y sólo si se satisfacen las siguientes tres condiciones:

- *Condición 1. El único entero positivo que divide (exactamente) tanto a c como a m es el uno.*
- *Condición 2. Si q es un número primo (es decir, es divisible únicamente por sí mismo y por uno) que divide a m , entonces q divide a $a - 1$.*
- *Condición 3. Si 4 divide a m , entonces 4 divide a $a - 1$.*

□

De este teorema se deduce que los generadores congruenciales lineales multiplicativos no pueden tener periodo completo, ya que no satisfacen la Condición 1: m es un entero positivo que divide tanto a m como a $c = 0$.

Generadores congruenciales lineales mixtos

Selección del módulo. El *módulo* del GCL mixto, m , suele escogerse de modo que dé lugar a periodo completo. Asimismo, interesa que el ciclo sea largo, con lo cual m es típicamente un número grande. Sin embargo, la dificultad que acarrea escoger m grande es que hace que la división por m , necesaria para generar los números z_i , sea una operación relativamente lenta.

Con el fin poder escoger m grande, y a la vez evitar tener que realizar explícitamente la división por m , se ha desarrollado una técnica denominada “*división por rebosamiento*”. Para poder aplicarla, el valor del módulo debe escogerse de la forma siguiente:

$$m = 2^b \quad (9.9)$$

donde b es el número de bits de la palabra binaria en la memoria el ordenador (descontando el bit de signo).

Por ejemplo, si el ordenador almacena los datos en palabras de 32 bits, y el más significativo se reserva para el signo, entonces: $b = 31 \rightarrow m = 2^{31} \approx 2.15E + 9$.

La elección $m = 2^b$ permite realizar la división por m sacando partido del rebosamiento: el mayor entero que puede representar una palabra de b bits es $2^b - 1$, con lo cual si se intenta almacenar un entero w mayor (con h bits, donde $h > b$), se producirá una pérdida de los $h - b$ bits más significativos. Los b bits que quedan almacenados son precisamente el resultado de la operación $w \pmod{2^b}$.

Ejemplo 9.4.2. *Se pretende programar en un ordenador "prehistorico" de 4 bits/palabra (sin considerar el bit de signo) el generador mostrado en el ejemplo 9.4.1.*

¿Cómo puede utilizarse el rebosamiento para obtener $z_7 = 12$ a partir de $z_6 = 5$? La representación binaria de $5 \cdot z_6 + 3 = 28$ es la palabra 11100. Como el ordenador sólo puede almacenar palabras de 4 bits, el bit más significativo se pierde, quedando almacenado 1100, que es la representación binaria de $z_7 = 12$. \square

Por los motivos expuestos anteriormente, aparentemente $m = 2^b$ es una buena elección para el módulo. Con esta elección, para que el generador mixto sea de periodo completo, a debe satisfacer las dos condiciones siguientes:

- Ser impar (Condición 1 del Teorema 9.4.1).
- $a - 1$ debe ser divisible por 4 (Condiciones 2 y 3 del Teorema 9.4.1).

Selección del multiplicador. En los primeros trabajos que aparecieron sobre generadores congruenciales lineales se prestó mucha atención a la eficiencia en el cálculo del producto $a \cdot z_{i-1}$. Esto condujo a expresiones del multiplicador de la forma:

$$a = 2^k + 1 \quad \text{donde } k \text{ es un entero positivo.} \quad (9.10)$$

El motivo de esta elección es que puede calcularse

$$a \cdot z_{i-1} = 2^k \cdot z_{i-1} + z_{i-1} \quad (9.11)$$

desplazando a la izquierda k bits la representación binaria de z_{i-1} , y sumando z_{i-1} al número obtenido del desplazamiento. No obstante, más adelante se descubrió que este tipo de generadores tiene propiedades estadísticas pobres, con lo cual no conviene usarlo.

¿Cómo puede entonces escogerse a y c , cuando $m = 2^b$, para obtener buenos generadores mixtos? Una respuesta indirecta a esta pregunta es que generalmente es preferible usar los generadores multiplicativos a los mixtos¹². Los generadores multiplicativos presentan similares prestación y son más sencillos de usar y de entender. Por este motivo, son ampliamente usados.

En cualquier caso, existen algunos generadores mixtos de probado buen comportamiento. Por ejemplo¹³:

$$\begin{aligned} z_i &= (5^{15} \cdot z_{i-1} + 1) \pmod{2^{35}} & \text{para } b = 35 \\ z_i &= (314159269 \cdot z_{i-1} + 453806245) \pmod{2^{31}} & \text{para } b = 31 \end{aligned} \quad (9.12)$$

Generadores congruenciales lineales multiplicativos

La gran ventaja de los generadores multiplicativos es su sencillez, ya que el valor de c queda fijado por definición: $c = 0$. Aparentemente, su desventaja es que no tienen periodo completo, ya que no satisfacen la Condición 1 del Teorema 9.4.1.

¹²Esta recomendación se recoge, entre otros, en el texto (Law & Kelton 2000).

¹³Referencia: (Law & Kelton 2000).

No obstante, en la práctica esta desventaja no supone ninguna limitación, ya que escogiendo m y a adecuadamente¹⁴ puede conseguirse un periodo de longitud $m - 1$. Las condiciones que deben satisfacerse en esta elección son las enunciadas en el Teorema 9.4.2. Los generadores multiplicativos en los que m y a se escogen de la forma indicada en dicho teorema se llaman *generadores congruenciales lineales multiplicativos de módulo primo*.

Teorema 9.4.2. *El periodo del generador congruencial lineal multiplicativo es igual a $m - 1$ si se cumplen las dos condiciones siguientes:*

- Condición 1. El módulo, m , es un número entero primo. Por ejemplo, igual al mayor número primo que sea menor que 2^b .
- Condición 2. El multiplicador, a , es una raíz primitiva de m . Es decir, el número entero más pequeño, k , para el cual $a^k - 1$ es divisible por m es: $k = m - 1$.

□

Los generadores congruenciales lineales multiplicativos de módulo primo tienen la propiedad de que se obtiene cada uno de los números $\{1, 2, \dots, m - 1\}$ una vez en cada ciclo (obsérvese que se ha excluido el cero), con lo cual la semilla puede ser cualquiera de estos números.

Surgen dos cuestiones acerca de este tipo de generadores:

1. ¿Cómo se obtiene una raíz primitiva del módulo (es decir, de m)? Desde el punto de vista computacional la tarea no es fácil, con lo cual habitualmente se recurre a valores ya calculados que se encuentran disponibles. Por ejemplo:

- Para $b = 35$. El mayor número primo que es menor que 2^{35} es: $m = 2^{35} - 31 = 34, 359, 738, 337$. Una raíz primitiva de $m = 2^{35} - 31$ es: $a = 5^5 = 3125$.
- Para $b = 31$. El mayor número primo que es menor que 2^{31} es¹⁵: $m = 2^{31} - 1 = 2, 147, 483, 647$. Dos posibles buenas elecciones de a son¹⁶:

$$a_1 = 7^5 = 16807 \quad \text{y} \quad a_2 = 630, 360, 016 \quad (9.13)$$

2. El valor de m se escoge de la forma: $m = 2^b - q$, donde q es un número entero positivo. Puesto que $m \neq 2^b$, no puede usarse el *desbordamiento entero* para calcular la división módulo m . En su lugar, se ha desarrollado una técnica denominada *división simulada*, que es descrita a continuación.

Técnica de la división simulada. El objetivo es calcular:

$$z_i = (a \cdot z_{i-1}) \pmod{2^b - q} \quad (9.14)$$

siendo conocido z_{i-1} . Para ello, en primer lugar se calcula z_i^* , que está definido de la forma siguiente:

$$z_i^* = (a \cdot z_{i-1}) \pmod{2^b} \quad (9.15)$$

Para calcular z_i^* se emplea la técnica del rebosamiento. A continuación, una vez conocido z_i^* , puede calcularse z_i de la expresión siguiente:

¹⁴Podría pensarse que escoger $m = 2^b$ podría resultar ventajoso, ya que evita realizar explícitamente la división. Sin embargo, puede demostrarse que con esta elección el periodo del generador multiplicativo es a lo sumo 2^{b-2} . Es decir, a lo sumo se generan un cuarto de los enteros $\{0, 1, \dots, m - 1\}$. Dado que no se sabe a priori dónde van a caer estos $\frac{m}{4}$ enteros, existe el riesgo de que formen grandes huecos en la secuencia $\{0, 1, \dots, m - 1\}$. Por este motivo, la elección $m = 2^b$ es desaconsejable en el caso de los generadores multiplicativos.

¹⁵Los números primos de la forma $2^p - 1$, donde p es un número entero, se llaman "números primos de Mersenne". Los primeros primos de Mersenne son 3, 7, 31, 127, etc. En la actualidad sólo se conocen 39 de ellos. El mayor de ellos fue encontrado a finales del año 2001. Se trata de $2^{13,466,917} - 1$, un número de más de 4 millones de dígitos. Puede encontrarse información interesante sobre este tema en: <http://www.mersenne.org/>

¹⁶El generador de números aleatorios que usaba Arena hasta fechas recientes era el siguiente generador congruencial lineal (Kelton et al. 2002):

$$m = 2^{31} - 1 = 2, 147, 483, 647 \quad , \quad a = 7^5 = 16807 \quad , \quad c = 0$$

El cual tiene una longitud de ciclo igual a $m - 1$. Este generador ha sido reemplazado (a partir de la versión 5.0 de Arena) por otro generador con mejores propiedades, tal como se discutirá más adelante.

$$z_i = \begin{cases} z_i^* + k \cdot q & \text{si } z_i^* + k \cdot q < 2^b - q \\ z_i^* + k \cdot q - (2^b - q) & \text{si } z_i^* + k \cdot q \geq 2^b - q \end{cases} \quad (9.16)$$

Donde k es el mayor número entero que es menor o igual que $\frac{a \cdot z_{i-1}}{2^b}$.

9.5. Otros generadores congruenciales

Aparte de los generadores congruenciales lineales, existen otros tipos de generadores desarrollados para obtener periodos más largos y mejores propiedades estadísticas. Sin embargo, un generador congruencial lineal con parámetros adecuados puede tener características tan buenas (y a veces mejores) que estas alternativas más complicadas.

Los generadores congruenciales lineales pueden considerarse un caso particular de un tipo más general de generadores, definidos mediante expresiones del tipo:

$$z_i = g(z_{i-1}, z_{i-2}, \dots) \pmod{m} \quad (9.17)$$

donde g es una función determinista. Como en el caso del generador lineal, los números obtenidos mediante esta fórmula son enteros comprendidos entre 0 y $m - 1$. A partir de ellos, se obtienen números distribuidos en el intervalo $[0, 1)$ mediante la transformación siguiente: $u_i = \frac{z_i}{m}$.

Los *generadores congruenciales cuadráticos* son una generalización obvia de los generadores lineales. Poseen la fórmula generatriz siguiente:

$$g(z_{i-1}, z_{i-2}, \dots) = a' \cdot z_{i-1}^2 + a \cdot z_{i-1} + c \quad (\text{Generador congruencial cuadrático}) \quad (9.18)$$

Puesto que z_i sólo depende de z_{i-1} (y no de números anteriores), y dado que $0 \leq z_i \leq m - 1$, en consecuencia el periodo de un generador congruencial cuadrático es a lo sumo m . El motivo es que en cuanto se genere un número que ya ha aparecido previamente, se repite la secuencia obtenida a partir de aquel. Dado que no hay más que m posibles números diferentes (los enteros de 0 a $m - 1$), el periodo es a lo sumo m .

Una elección diferente de la función g es mantener la linealidad, pero usando más valores previos. Esto da lugar a generadores de la forma:

$$g(z_{i-1}, z_{i-2}, \dots) = a_1 \cdot z_{i-1} + a_2 \cdot z_{i-2} + \dots + a_q \cdot z_q \quad (9.19)$$

donde a_1, \dots, a_q son constantes. Escogiendo los parámetros adecuadamente pueden conseguirse periodos tan grandes como $m^q - 1$.

Un caso particular de este tipo de generadores es el *generador de Fibonacci* (el cual posee malas propiedades estadísticas y, por tanto, es desaconsejable su uso):

$$z_i = (z_{i-1} + z_{i-2}) \pmod{m} \quad (\text{Generador de Fibonacci}) \quad (9.20)$$

Otra línea para la generalización de los generadores congruenciales lineales ha consistido en modificar el multiplicador (a) y el incremento (c) de acuerdo con fórmulas congruenciales antes de generar el nuevo z_i . Estos generadores permiten obtener periodos enormes.

9.6. Generadores combinados

Algunos investigadores han propuesto métodos para combinar dos o más generadores independientes (generalmente dos o tres generadores congruenciales lineales) de modo que el

generador resultante tenga mejores propiedades estadísticas que cada uno de los generadores independientes que lo componen. La desventaja de este método es que aumenta el coste computacional de generar cada número pseudo aleatorio.

Un generador combinado bien conocido, que emplea dos generadores congruenciales lineales, es el siguiente:

1. Inicialmente se llena un vector $\mathbf{V} = (v_1, v_2, \dots, v_k)$ secuencialmente con los primeros k números obtenidos del primer generador congruencial lineal. Se sugiere emplear $k = 128$.
2. El segundo generador congruencial lineal se usa para obtener un entero I distribuido uniformemente entre los enteros $\{1, 2, \dots, k\}$. Este número entero I es el índice del elemento del vector \mathbf{V} que se toma como primer número pseudo aleatorio obtenido por el generador compuesto: v_I .
3. A continuación se genera un nuevo número para reemplazar a v_I en el vector \mathbf{V} . Para ello se usa el primer generador.
4. De nuevo, se usa el segundo generador para obtener el índice del elemento de \mathbf{V} que será el siguiente número obtenido por el generador compuesto. Y así sucesivamente.

Este método, que emplea un segundo generador para muestrear una secuencia obtenida por un primer generador, es de esperar que rompa cualquier tipo de correlación que exista en las secuencias de los generadores individuales y que también extienda considerablemente el periodo. Se ha demostrado que es posible obtener generadores combinados con muy buenas propiedades estadísticas usando generadores congruenciales lineales con propiedades estadísticas pobres. Sin embargo, se ha demostrado que se obtiene poca mejora cuando se usan buenos generadores congruenciales lineales.

Asimismo, se ha demostrado que se obtienen resultados igualmente buenos usando vectores de longitud $k = 2$ a los obtenidos usando vectores considerablemente más largos.

Se han propuesto otros muchos tipos de generadores combinados: usar un único generador para muestrear la secuencia obtenida de él mismo, aplicar permutaciones fijas a la secuencia (con lo que sólo es preciso usar un generador), etc.

Un problema de este tipo de generadores es que no es posible realizar saltos en la secuencia de números sin generar todos los valores intermedios, cosa que sí puede hacerse con los generadores congruenciales lineales.

Para ilustrar la finalidad de dar saltos en la secuencia, considérese que se desean obtener varias secuencias, de 1,000,000 de números cada una, para simular diferentes fenómenos aleatorios (se usa una secuencia para obtener valores de cada fenómeno: intervalos entre llegadas, tiempos de proceso, intervalos de tiempo entre averías, etc.). Con un generador congruencial lineal, puede usarse como semilla para la primera secuencia z_0 , para la segunda secuencia $z_{1,000,000}$, para la tercera $z_{2,000,000}$, y así sucesivamente.

Existen formas de combinar varios generadores que no presentan esta dificultad. Por ejemplo, si $\{z_{1,i}\}$ y $\{z_{2,i}\}$ representan las secuencias de números enteros obtenidas de dos generadores congruenciales lineales con diferente módulo, la secuencia de números del generador compuesto puede calcularse de la forma siguiente:

$$z_i = (z_{1,i} - z_{2,i}) \pmod{m} \quad (9.21)$$

Esta idea puede extenderse a más de dos generadores y tiene varias ventajas:

- el periodo puede ser muy grande (al menos tan grande como 10^{18}),
- los multiplicadores de cada uno de los generadores pueden ser pequeños,
- el generador compuesto es rápido, y
- tiene buenas propiedades estadísticas.

Generador de números aleatorios de Arena

El generador de números aleatorios instalado en la versión 5.0 y posteriores de Arena¹⁷ es del tipo denominado “*generadores recursivos combinados múltiples*”. Tiene un ciclo de longitud $3.1 \cdot 10^{57}$. Para iniciar la recursión, hace falta una semilla, que es un vector de 6 componentes: $\{A_0, A_1, A_2, B_0, B_1, B_2\}$. Opera de la manera siguiente:

1. Comienza con las siguientes dos fórmulas recursivas separadas, que puede considerarse que son operadas en paralelo:

$$A_n = (1403580 \cdot A_{n-2} - 810728 \cdot A_{n-3}) \pmod{4294967087} \quad (9.22)$$

$$B_n = (527612 \cdot B_{n-1} - 1370589 \cdot B_{n-3}) \pmod{4294944443} \quad (9.23)$$

2. A continuación, se combinan los dos valores obtenidos anteriormente de la forma siguiente:

$$Z_n = (A_n - B_n) \pmod{4294967087} \quad (9.24)$$

3. Finalmente, se calcula el número pseudo aleatorio de la manera siguiente:

$$U_n = \begin{cases} \frac{Z_n}{4294967088} & \text{si } Z_n > 0 \\ \frac{4294967087 - Z_n}{4294967088} & \text{si } Z_n = 0 \end{cases} \quad (9.25)$$

Resulta útil separar el ciclo del generador en secuencias adyacentes, no solapadas. Para definir estas secuencias es necesario definir vectores semilla lo suficientemente espaciados entre sí. El generador de Arena permite, de forma sencilla, separar su ciclo, de $3.1 \cdot 10^{57}$ números, en $1.8 \cdot 10^{19}$ secuencias, cada una de ellas de longitud $1.7 \cdot 10^{38}$.

Cada una de estas secuencias es, a su vez, dividida en $2.3 \cdot 10^{15}$ sub-secuencias, cada una de ellas con una longitud $7.6 \cdot 10^{22}$. Como ilustración de lo inmensa que es la longitud de cada una de estas sub-secuencias, basta considerar que un PC a 600 MHz tardaría 1.8 billones de años en agotar la sub-secuencia, realizando únicamente la operación de generar el número pseudo aleatorio.

En Arena, cuando se generan observaciones a partir de una distribución, puede especificarse qué secuencia de números aleatorios se desea emplear para ello. La forma de hacerlo consiste en añadir el número de secuencia a continuación de los parámetros de la distribución. Por ejemplo, para generar observaciones de una distribución exponencial con media 6.7, usando la secuencia 4, debe escribirse: *EXPO(6.7, 4)*.

Si no se especifica el número de la secuencia, por defecto Arena usa la número 10. Asimismo, Arena usa la secuencia 10 para generar los números aleatorios necesarios para los procesos de llegada no estacionarios o los módulos *Decide*. Por este motivo, cuando el usuario escoge una secuencia, conviene que no sea la número 10.

La idea de usar secuencias de números aleatorios diferentes para diferentes propósitos en la simulación (por ejemplo, la secuencia 1 para los tiempos entre llegadas de las entidades, la secuencia 2 para el tipo de las entidades, la secuencia 3 para los tiempos de proceso, etc.) será discutida al exponer las *técnicas de reducción de la varianza*.

Cuando se realizan múltiples réplicas de la simulación del modelo, Arena salta automáticamente, con cada nueva réplica, al comienzo de la siguiente sub-secuencia dentro de cada secuencia, de modo que cada réplica se realice con una sub-secuencia de números aleatorios diferentes.

9.7. Generadores de Tausworthe

Este tipo de generadores opera directamente con los bits para formar números pseudo aleatorios. Se define una secuencia b_1, b_2, \dots de dígitos binarios (bits) mediante una fórmula del tipo siguiente:

¹⁷El contenido de este apartado está extraído del texto (Kelton et al. 2002).

$$b_i = (c_1 \cdot b_{i-1} + c_2 \cdot b_{i-2} + \dots + c_q \cdot b_{i-q}) \pmod{2} \quad (9.26)$$

donde c_1, c_2, \dots, c_q son constantes que pueden tomar el valor 0 ó 1.

El periodo máximo del generador puede obtenerse a partir de los razonamientos siguientes:

- El valor del bit generado depende de los q bits anteriores de la secuencia. Con lo cual, una vez que los q bits anteriores tomen un valor que ya ha aparecido anteriormente, se repite la secuencia.
- Con q bits se pueden formar 2^q palabras diferentes.
- La palabra con todos los bits a cero debe excluirse del ciclo, ya que una vez que aparece, los sucesivos bits generados son cero.

En consecuencia, el periodo máximo de este tipo de generadores es igual a $2^q - 1$. Es decir, el número de palabras posibles menos la palabra cero.

En la inmensa mayoría de las aplicaciones sólo dos de los coeficientes son diferentes de cero. En estos casos, la fórmula del generador es la siguiente:

$$b_i = (b_{i-r} + b_{i-q}) \pmod{2} \quad \text{donde } r, q \text{ son enteros, tales que } 0 < r < q \quad (9.27)$$

Para inicializar la secuencia es preciso proporcionar los primeros q bits: b_1, \dots, b_q , que constituyen la semilla.

Ejemplo 9.7.1. *Dados los valores: $r = 3$, $q = 5$ y $b_1 = b_2 = b_3 = b_4 = b_5 = 1$, los primeros 42 bits de la secuencia son:*

1 1 1 1 1 0 0 0 1 1 0 1 1 1 0 1 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0
1 1 1 1 1 0 0 0 1 1 0

Obsérvese que el periodo de este generador es $2^q - 1 = 31$. □

Existen diferentes métodos de transformar la secuencia de bits, $\{b_i\}$, en números distribuidos en el intervalo $[0, 1)$. Dos de estos métodos son los siguientes:

- Un método consiste en agrupar las cadenas de i bits consecutivos para formar el entero binario de i bits entre 0 y $2^i - 1$, que es entonces dividido por 2^i . En este caso, i no debe ser superior al tamaño de palabra del ordenador.

Ejemplo 9.7.2. *En la secuencia del Ejemplo 9.7.1, escogiendo $i = 4$, resulta la secuencia siguiente:*

15/16, 8/16, 13/16, 13/16, 4/16, 2/16, 5/16, 9/16, 15/16, 1/16, ...

□

- Otra posibilidad es usar los i primeros bits consecutivos de la secuencia para obtener el primer entero, saltar cierto número de los bits siguientes, obtener el siguiente entero de los siguientes i bits, volver a saltar, y así sucesivamente.

Los generadores de Tausworthe ofrecen ciertas ventajas sobre los generadores congruenciales lineales:

- Son esencialmente independientes del tamaño de a palabra del ordenador.
- Pueden obtenerse periodos inimaginablemente largos (tales como $2^{521} - 1 > 10^{156}$).

El ingrediente fundamental de los generadores de Tausworthe es una secuencia de bits, que puede ser generada de diferentes formas. Una forma alternativa a la expuesta anteriormente de generar la secuencia de bits es la siguiente¹⁸:

1. Sean p y q números primos grandes, con $p - 3$ y $q - 3$ ambos divisibles por 4.

¹⁸Aparece descrita en el texto (Law & Kelton 2000).

2. Los números p y q son usados para calcular un módulo, $m = p \cdot q$, que se usa para obtener una secuencia de números enteros usando la fórmula siguiente:

$$x_i = x_{i-1}^2 \pmod{m} \quad (9.28)$$

3. Se define la secuencia de bits de la forma siguiente:

$$b_i = \text{bit menos significativo de } x_i \quad (9.29)$$

Esta secuencia de bits es prácticamente impredecible, ya que descubrir la no aleatoriedad (es decir, descubrir de forma precisa cuál es el valor de un bit que falta) es computacionalmente equivalente a factorizar m en $p \cdot q$, un problema que requiere ingentes recursos de computación.

9.8. Tests empíricos

Posiblemente, la manera más sencilla de analizar un generador es usarlo para obtener una secuencia de números, y a continuación estudiar con qué probabilidad estos números podrían haber sido obtenidos muestreando una distribución $U(0, 1)$. Una consecuencia evidente es que, puesto que sólo se examina un segmento del ciclo, no se puede afirmar nada acerca del resto del ciclo, que no ha sido estudiado. Existen innumerables tests de este tipo¹⁹, los cuales se denominan genéricamente *tests empíricos*.

Existen algunas pruebas sencillas que, si bien no constituyen tests estadísticos, pueden proporcionar cierta confianza acerca de la calidad del generador, o bien despertar serias dudas. Por ejemplo, cuando se genera una larga secuencia de números pseudo aleatorios, pueden obtenerse a partir de ella estimaciones muy precisas de la media, la varianza y el factor de autocorrelación, que deberían estar muy próximas a los valores teóricos, que son:

$$E\{U\} = \frac{1}{2} \quad (9.30)$$

$$\text{Var}\{U\} = \frac{1}{12} \quad (9.31)$$

$$E\{U_i \cdot U_{i+1}\} = E\{U_i\} \cdot E\{U_{i+1}\} = \frac{1}{4} \quad (9.32)$$

Ejemplo 9.8.1. *El siguiente generador fue propuesto por Grogono para ordenadores de 16 bits:*

$$z_0 = 23311 \quad (9.33)$$

$$z_i = (25173 \cdot z_{i-1} + 13849) \pmod{65536} \quad (9.34)$$

Si se generan 10.000 números y se normalizan al intervalo unidad, se obtienen los estimadores siguientes:

$$\bar{Z}(10.000) = 0.5019 \quad (9.35)$$

$$S_Z^2(10.000) = 0.08412 \quad (9.36)$$

$$\frac{1}{10.000} \cdot \sum_{i=1}^{9999} z_i \cdot z_{i+1} = 0.2526 \quad (9.37)$$

los cuales están muy próximos a los valores teóricos. □

Las dos propiedades deseables de los números aleatorios son la *uniformidad* y la *independencia*. Para contrastar si una determinada secuencia de números las posee, se han desarrollado una serie de tests, que pueden clasificarse en dos categorías, de acuerdo con estas dos propiedades de interés: *tests de uniformidad* y *tests de independencia*.

¹⁹Ver, por ejemplo, el texto (Law & Kelton 2000).

- *Tests de uniformidad.* La hipótesis nula es que los números están distribuidos uniformemente en el intervalo $[0, 1]$.
- *Tests de independencia.* La hipótesis nula es que los números de la secuencia son independientes entre sí.

Cuando al aplicar un test de uniformidad o un test de independencia se falla en refutar la hipótesis nula, significa que no se ha encontrado evidencia de uniformidad o dependencia empleando ese determinado test en concreto. No obstante, siempre queda abierta la posibilidad de que se descubra desuniformidad o dependencia usando un test diferente.

Como ya se adelantó al comienzo del tema, la dificultad fundamental que debe afrontarse en la búsqueda de la aleatoriedad es que mientras existen muchas formas de probar que una secuencia no es aleatoria, no existe forma alguna de probar que sí lo es.

Baterías de tests

Un procedimiento para analizar mediante test empíricos un generador de números pseudo aleatorios consiste en obtener a partir de él un determinado número de secuencias de números y contrastar la uniformidad e independencia de cada una de ellas. Para ello, generalmente no se aplica un único test de uniformidad y un único test de independencia, sino que se aplica un conjunto de ellos, es decir, una *batería*²⁰ de tests. Al analizar los resultados obtenidos, deben tenerse en consideración los hechos siguientes.

Al aplicar un test, debe establecerse el valor de su nivel de significación, α . El *nivel de significación del test* es la probabilidad de que el test rechace la hipótesis nula si ésta es verdadera. Es decir:

$$\alpha = \text{Prob} \{ \text{rechazar } H_0 \mid H_0 \text{ es verdadera} \} \quad (9.38)$$

Frecuentemente se escoge el nivel de significación igual a 0.01 ó a 0.05.

Si se realizan varios tests sobre una misma secuencia de números, la probabilidad de rechazar la hipótesis nula en al menos un test, únicamente por azar, aumenta. En concreto, la probabilidad de que en uno de N test rechace la hipótesis nula, siendo ésta verdadera, es igual a $\sum_{i=1}^N \alpha_i$. Este hecho se ilustra en el siguiente ejemplo.

Ejemplo 9.8.2. *Supóngase que se realizan 5 tests sobre una misma secuencia de números, y que el nivel de significación de cada uno de los tests es $\alpha = 0.05$. Entonces, la probabilidad de rechazar la hipótesis nula en uno de los tests, únicamente por azar, es igual a 0.25 (es decir, el 25 %).* □

Similarmente, si se realiza un test repetidamente sobre varias secuencias de un mismo generador, la probabilidad de rechazar la hipótesis nula en al menos uno de estos tests, únicamente por azar, aumenta al aumentar el número de secuencias testeadas.

Ejemplo 9.8.3. *Si se analizan 100 secuencias diferentes de un mismo generador, aplicando a cada una de ellas un test con un nivel de significación $\alpha = 0.05$, entonces cabe esperar que aproximadamente 5 de estos tests fallen únicamente por azar. En consecuencia, si el número de secuencias en las que el test rechaza la hipótesis nula está próximo a $100 \cdot \alpha = 5$, entonces ésta no es razón suficiente para descartar el generador.* □

El análisis de los resultados de aplicar una batería de tests a un conjunto de secuencias de números pseudo aleatorios, puede realizarse siguiendo la metodología descrita a continuación²¹.

²⁰Pueden encontrarse en internet programas con baterías de test empíricos que pueden ser aplicadas para contrastar la validez de los generadores. Entre éstas destaca *Diehard*, una batería de tests desarrollada por G. Marsaglia. Puede encontrarse información en la dirección <http://stat.fsu.edu/geo/diehard.html>

²¹Extraída del texto (Guasch, Piera, Casanovas & Figueras 2002).

| | Serie 1 | Serie 2 | Serie 3 | Serie 4 | Serie 5 | Serie 6 | Serie 7 | Serie 8 | Serie 9 | Serie 10 |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| Test 1 | S | S | S | S | S | S | S | S | N | S |
| Test 2 | S | S | S | S | S | S | S | S | S | S |
| Test 3 | S | N | S | S | S | S | S | S | S | S |
| Test 4 | S | S | S | S | S | S | S | S | S | S |
| Test 5 | S | S | S | N | S | S | S | S | S | S |
| Test 6 | S | S | S | S | S | S | S | N | S | S |
| Test 7 | S | S | S | S | S | S | S | S | S | S |
| Test 8 | S | S | S | S | S | S | S | S | S | S |

Caso A: El generador pasa la batería de tests.

| | Serie 1 | Serie 2 | Serie 3 | Serie 4 | Serie 5 | Serie 6 | Serie 7 | Serie 8 | Serie 9 | Serie 10 |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| Test 1 | S | S | S | N | S | S | S | S | N | S |
| Test 2 | S | S | S | N | S | S | N | S | S | S |
| Test 3 | S | N | S | N | S | S | N | S | S | S |
| Test 4 | S | S | S | S | S | S | S | S | S | S |
| Test 5 | N | N | N | N | S | N | N | N | N | S |
| Test 6 | S | S | S | N | S | S | N | N | S | S |
| Test 7 | N | N | N | N | N | S | N | N | N | N |
| Test 8 | S | S | S | S | S | S | N | S | S | S |

Caso B: El generador no pasa la batería de tests.

Figura 9.6: Ejemplo: batería de tests aplicada a secuencias de dos generadores.

La metodología consiste en construir una tabla de doble entrada, en la que cada fila especifican el test aplicado y las columnas se etiquetan con el identificador de cada secuencia. La tabla se rellena con las respuestas de los distintos tests, indicando si la secuencia pasa el test (S), o si no pasa el test (N). Se considera aceptable que una tabla tenga “unos pocos” elementos de resultado negativo. Por ejemplo, no serían aceptables generadores tales que todas las secuencias fallaran un determinado test, o tales que una secuencia falle todos los tests.

Ejemplo 9.8.4. En la Figura 9.6 se muestran las tablas obtenidas de aplicar una batería de 8 tests empíricos a dos generadores. Los tests se han aplicado a 10 secuencias de números obtenidas de cada uno de los generadores. Como puede observarse en la Figura 9.6, el primero de los generadores pasa la batería de tests, mientras que el segundo no la pasa. \square

9.9. Tests empíricos de uniformidad

Los *tests de uniformidad* (también denominados *tests de la frecuencia*) comparan la distribución de la secuencia de números, x_1, \dots, x_N , con la distribución uniforme. La hipótesis nula es que no existe diferencia entre ambas distribuciones.

Tests de uniformidad ampliamente usados son el *test de Kolmogorov-Smirnov* y el *test chi-cuadrado*. El *test serie* (“serial test” en lengua inglesa) es una generalización del test chi-cuadrado a dimensiones superiores a uno. Agrupando cada d números consecutivos de la secuencia, se define un punto en el espacio d -dimensional por cada uno de estos grupos. El

test contrasta si estos puntos están uniformemente distribuidos en un cubo d -dimensional de lado uno.

Test de Kolmogorov-Smirnov

El *test de Kolmogorov-Smirnov* (abreviadamente, *test K-S*) compara la probabilidad acumulada de la distribución uniforme, $F_X(x) = x$ para $0 \leq x \leq 1$, con la probabilidad acumulada empírica obtenida de la secuencia, $F_X^*(x) = \frac{\text{número de } x_1, \dots, x_N \text{ que son } \leq x}{N}$.

Si la hipótesis nula es verdadera, a medida que se aumenta la longitud de la secuencia, N , la distribución empírica $F_X^*(x)$ se va aproximando cada vez más a la distribución teórica $F_X(x)$. El *test de Kolmogorov-Smirnov* se basa en la mayor desviación absoluta entre $F_X^*(x)$ y $F_X(x)$ sobre el rango de la variable aleatoria.

El estadístico del test es el siguiente:

$$D = \text{máx} |F_X(x) - F_X^*(x)| \tag{9.39}$$

Si se satisface la hipótesis nula, el valor obtenido del estadístico D es una observación aleatoria de una distribución que es conocida (está tabulada: ver la Tabla T.3, al final del libro), y que depende de la longitud, N , de la secuencia.

Para aplicar el test de K-S a la comparación de la distribución de la secuencia con la distribución uniforme, pueden seguirse los pasos siguientes:

1. Ordenar los números de la secuencia en orden creciente. Sea $x_{(i)}$ la i -ésima menor observación. Es decir: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$.
2. Calcular la distancia máxima, D , entre las dos probabilidades acumuladas, de la manera siguiente:

$$D^+ = \text{máx}_{1 \leq i \leq N} \left\{ \frac{i}{N} - x_{(i)} \right\} \tag{9.40}$$

$$D^- = \text{máx}_{1 \leq i \leq N} \left\{ x_{(i)} - \frac{i-1}{N} \right\} \tag{9.41}$$

$$D = \text{máx} \{D^+, D^-\} \tag{9.42}$$

Este procedimiento para calcular D se basa en que la distancia vertical máxima entre las dos gráficas siempre se da en uno de los puntos en los que se produce un salto en la distribución empírica. Por este motivo, sólo es preciso considerar aquellos valores del intervalo $[0, 1]$ que corresponden con alguno de los valores de la secuencia.

3. Determinar el valor crítico, D_α , de la Tabla T.3, para el nivel de significación deseado, α , y la longitud de la secuencia, N .
4. Si el estadístico de la secuencia D es mayor que el valor crítico, D_α , entonces se rechaza la hipótesis nula. Por el contrario, si $D \leq D_\alpha$, entonces no existe evidencia para rechazar la hipótesis nula, con lo cual la secuencia pasa el test de K-S.

Ejemplo 9.9.1. Se desea aplicar el test de Kolmogorov-Smirnov con un nivel de significación $\alpha = 0.05$ a la secuencia de cinco números siguiente:

0.44 0.81 0.14 0.05 0.93

La presentación en una tabla de los cálculos intermedios facilita su comprensión:

| | | | | | |
|---------------------------|------|------|------|------|------|
| $x_{(i)}$ | 0.05 | 0.14 | 0.44 | 0.81 | 0.93 |
| $\frac{i}{N}$ | 0.20 | 0.40 | 0.60 | 0.80 | 1.00 |
| $\frac{i}{N} - x_{(i)}$ | 0.15 | 0.26 | 0.16 | - | 0.07 |
| $x_{(i)} - \frac{i-1}{N}$ | 0.05 | - | 0.04 | 0.21 | 0.13 |

Los estadísticos valen $D^+ = 0.26$ y $D^- = 0.21$, con lo cual, $D = 0.26$.

El valor crítico de D , obtenido de la Tabla T.3, para $\alpha = 0.05$ y $N = 5$, es 0.565. Puesto que el valor calculado, 0.26, es menor que el valor tabulado, 0.565, no se rechaza la hipótesis de que la distribución de los números generados es igual a la distribución uniforme. \square

Test chi-cuadrado

El test chi-cuadrado también puede aplicarse para contrastar la hipótesis de que la distribución empírica, obtenida de la secuencia de números, está distribuida uniformemente en el intervalo unidad. Para ello, debe tenerse en cuenta que no se han usado los datos experimentales (la secuencia de números) para ajustar ninguno de los parámetros de la distribución.

La aplicación del test a una secuencia de números u_1, \dots, u_n consiste en los pasos siguientes:

1. Se divide el intervalo $[0, 1]$ en k subintervalos (denominados *clases*) de igual longitud²². Sea f_j el número de datos de la secuencia que pertenecen a la clase j , es decir, que se encuentran en el intervalo j -ésimo.
2. Calcular el valor del siguiente estadístico, que proporciona una medida de la diferencia entre la frecuencia observada en la secuencia, f_j , y la frecuencia que cabría esperar si los datos estuvieran distribuidos de forma perfectamente uniforme, $\frac{n}{k}$:

$$\chi^2 = \frac{k}{n} \cdot \sum_{j:1}^k \left(f_j - \frac{n}{k} \right)^2 \quad (9.43)$$

Un valor pequeño del estadístico χ^2 indica una buena coincidencia entre las frecuencias teóricas y las experimentales.

3. Si la longitud de la secuencia, n , es grande y si se satisface la hipótesis nula:

H_0 : las observaciones u_1, \dots, u_n están distribuidas $U(0, 1)$

entonces el estadístico χ^2 es una observación de la distribución chi-cuadrado con $k - 1$ grados de libertad.

Por consiguiente, si $\chi^2 > \chi_{k-1, 1-\alpha}^2$, entonces el test rechaza la hipótesis nula con nivel α .

Por el contrario, si $\chi^2 \leq \chi_{k-1, 1-\alpha}^2$, entonces no existe evidencia suficiente para rechazar la hipótesis nula.

Frecuentemente, se emplea un número de clases muy elevado. En estos casos, puede usarse la siguiente aproximación para calcular el punto crítico de la distribución chi-cuadrado con $k - 1$ grados de libertad²³:

$$\chi_{k-1, 1-\alpha}^2 \approx (k-1) \cdot \left\{ 1 - \frac{2}{9 \cdot (k-1)} + z_{1-\alpha} \cdot \sqrt{\frac{2}{9 \cdot (k-1)}} \right\}^3 \quad (9.44)$$

donde $\chi_{k-1, 1-\alpha}^2$ es punto crítico $1 - \alpha$ de la distribución chi-cuadrado con $k - 1$ grados de libertad, y $z_{1-\alpha}$ es punto crítico $1 - \alpha$ de la distribución $N(0, 1)$.

Ejemplo 9.9.2. Se desea estudiar, mediante la aplicación del test chi-cuadrado, la uniformidad de una secuencia obtenida de un generador de un solo dígito. La secuencia consta de $n = 250$ dígitos. Cada dígito puede tomar un valor entero comprendido entre 0 y 9.

Se agrupan los dígitos de la secuencia en $k = 10$ (una clase por cada dígito), obteniéndose la siguiente frecuencia experimental para cada clase:

²²En el texto (Law & Kelton 2000) se ofrece la siguiente recomendación:

- el número de clases, k , debe ser al menos 100, y
- $\frac{n}{k}$ debe ser al menos 5.

²³(Law & Kelton 2000).

| | | | | | | | | | | |
|--------|----|----|----|----|----|----|----|----|----|----|
| Dígito | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| f_j | 30 | 20 | 35 | 36 | 17 | 14 | 29 | 20 | 18 | 31 |

El estadístico vale $\chi^2 = \frac{1}{25} \sum_{i=1}^{10} (f_i - 25)^2 = 23.29$. Para un nivel de significación del 5% (es decir, $\alpha = 0.05$), el valor del punto crítico $\chi_{9,0.95}^2 = 16.919$ (ver Tabla T.2, al final del libro). Puesto que $\chi^2 = 23.29 > \chi_{9,0.95}^2 = 16.919$, puede concluirse que este generador no se comportan, con un nivel de significación del 5%, como un generador de variables aleatorias independientes e idénticamente distribuidas. \square

Los tests chi-cuadrado y K-S pueden aplicarse indistintamente para contrastar la uniformidad de la muestras grandes de datos. Sin embargo, si la longitud de la secuencia es pequeña²⁴, entonces debe aplicarse el test K-S y no el test chi-cuadrado.

9.10. Tests empíricos de independencia

En la sección anterior se han discutido dos test para contrastar la uniformidad de la secuencia: el test K-S y el test chi-cuadrado. Además de la uniformidad, es preciso contrastar la independencia de la secuencia.

Con el fin de ilustrar este hecho, supóngase una secuencia de 100 números tales que los 10 primeros están en el rango 0.01 – 0.1, los 10 siguientes están en el rango 0.11 – 0.20, y así sucesivamente. Esta secuencia pasará los tests de frecuencia con facilidad, pero la ordenación de los números producida por el generador no es aleatoria.

Los test ideados para descubrir las dependencias entre los números de la secuencia se denominan *tests de independencia*. Existen muchos tipos diferentes de tests de independencia. En este texto únicamente se discutirán los cuatro siguientes: los *tests de las carreras*, el *test de autocorrelación*, el *test de los huecos* y el *test de poker*.

Tests de las carreras

Este tipo de tests de independencia hacen uso del concepto de “*carrera*”. Una *carrera* se define como

“una sucesión de eventos iguales, antecedida y sucedida por un evento diferente”.

La *longitud de una carrera* es el número de eventos que la componen. A continuación se muestra un ejemplo.

Ejemplo 9.10.1. En un experimento consistente en el lanzamiento de una moneda, cada lanzamiento tiene los 3 posibles resultados siguientes:

- que se obtenga cara,
- que se obtenga cruz, o
- que no se produzca el evento (evento nulo). Se considera que cada secuencia de lanzamientos comienza y termina con un evento nulo.

Considérese la siguiente secuencia de lanzamientos de una moneda, donde “+” representa “cara” y “-” representa “cruz”:

+ - - + + - - - + -

²⁴En el texto (Banks et al. 1996) se indica que, en este contexto, la secuencia se considera “pequeña” cuando está compuesta por menos de 50 números.

Esta secuencia de 10 lanzamientos contiene las 6 carreras siguientes:

$$\{+\} \quad \{-, -\} \quad \{+, +\} \quad \{-, -, -\} \quad \{+\} \quad \{-\}$$

cuyas longitudes son respectivamente: 1, 2, 2, 3, 1 y 1. □

Existen varios tipos diferentes de tests de las carreras. En este texto únicamente se discutirán los tres siguientes:

- Test de las carreras crecientes y decrecientes.
- Test de las carreras por encima y por debajo de la media.
- Test de la longitud de las carreras.

Test de las carreras crecientes y decrecientes

Una *carrera creciente* es una secuencia en la cual cada número está seguido de uno mayor que él. Similarmente, una *carrera decreciente* es una secuencia en la cual cada número está seguido por otro menor que él. Para ilustrar estos conceptos, considérese el siguiente ejemplo.

Ejemplo 9.10.2. En la secuencia de 15 números siguiente:

$$0.57 \quad 0.10 \quad 0.33 \quad 0.42 \quad 0.59 \quad 0.21 \quad 0.20 \quad 0.09 \\ 0.24 \quad 0.15 \quad 0.73 \quad 0.89 \quad 0.98 \quad 0.12 \quad 0.74$$

si se sustituye cada número por “+” o “-”, dependiendo de que el siguiente número sea mayor o menor que él respectivamente, se obtiene:

$$-, +, +, +, -, -, -, +, -, +, +, +, -, +$$

Cada sucesión de símbolos “+” consecutivos, o cada sucesión de símbolos “-” consecutivos, forma una carrera. En esta secuencia hay 8 carreras:

$$\{-\}, \{+, +, +\}, \{-, -, -\}, \{+\}, \{-\}, \{+, +, +\}, \{-\}, \{+\}$$

de las cuales 4 son crecientes (secuencias de símbolos “+”) y 4 son decrecientes (secuencias de símbolos “-”). La longitud de las 8 carreras es 1, 3, 3, 1, 1, 3, 1 y 1 respectivamente. □

Tan improbable es que una secuencia tenga muchas carreras como que tenga pocas. Por ejemplo, la secuencia de números:

$$0.02 \quad 0.28 \quad 0.33 \quad 0.35 \quad 0.43 \quad 0.51 \quad 0.61 \quad 0.81 \quad 0.89 \quad 0.98$$

tiene una única carrera, que es creciente. Es muy poco probable que una secuencia aleatoria de 10 números tenga una única carrera. Asimismo, la secuencia:

$$0.08 \quad 0.93 \quad 0.15 \quad 0.96 \quad 0.26 \quad 0.84 \quad 0.28 \quad 0.79 \quad 0.36 \quad 0.57$$

tiene 9 carreras: 5 crecientes y 4 decrecientes. También es muy improbable que una secuencia aleatoria de 10 números tenga 9 carreras. Intuitivamente se comprende que lo más probable es que el número de carreras esté “entre los dos extremos”.

En una secuencia aleatoria de N números, el número de carreras es una variable aleatoria, A , cuya media y varianza son las siguientes²⁵:

$$\mu_A = \frac{2 \cdot N - 1}{3} \tag{9.45}$$

$$\sigma_A^2 = \frac{16 \cdot N - 29}{90} \tag{9.46}$$

²⁵(Banks et al. 1996).

y supóngase que prosigue de la misma manera: dos números por encima de la media, dos números por debajo de la media y así sucesivamente. El test de las carreras por encima y por debajo de la media no detectaría esta falta de independencia. Sin embargo, cabría esperar que una secuencia independiente no sólo contuviera carreras de longitud 2.

El test de longitud de las carreras consiste en contrastar la diferencia, entre

- el valor, observado en la secuencia, del número de carreras de cada longitud, y
- el valor esperado, en una secuencia independiente, del número de carreras de cada longitud.

La comparación se establece aplicando el test chi-cuadrado, donde las diferentes longitudes de las carreras son las clases. El test puede aplicarse tanto a carreras crecientes y decrecientes, como a carreras por encima y por debajo de la media. Si bien, en cada caso el número esperado de carreras de cada longitud es diferente²⁹.

Test de autocorrelación

Otro tipo de test empírico que contrasta la independencia de la secuencia es el *test de autocorrelación*. Para ilustrar qué tipo de dependencias identifica este tipo de test, considérese la siguiente secuencia de números:

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 0.12 | 0.01 | 0.23 | 0.28 | 0.89 | 0.31 | 0.64 | 0.28 | 0.83 | 0.93 |
| 0.99 | 0.15 | 0.33 | 0.35 | 0.91 | 0.41 | 0.60 | 0.27 | 0.75 | 0.88 |
| 0.68 | 0.49 | 0.05 | 0.43 | 0.95 | 0.58 | 0.19 | 0.36 | 0.69 | 0.87 |

Inspeccionando visualmente los números, parecen aleatorios, y probablemente pasarían todos los tests presentados hasta este punto. Sin embargo, examinando las posiciones 5, 10, 15, etc., se observa que hay números muy grandes en esa posición. En efecto, 30 números es una muestra muy pequeña para descartar un generador, pero la idea es que los números que ocupan unas determinadas posiciones en la secuencia pueden estar relacionados entre sí.

El test de autocorrelación está indicado para contrastar si este tipo de relación existe. La relación puede ser, o bien que los números de determinadas posiciones sean todos “grandes”, que sean todos “pequeños”, o que vayan alternándose números grandes y pequeños.

Considérese la secuencia de números siguiente:

$$x_1, \dots, x_N \quad (9.64)$$

y la siguiente sub-secuencia de la misma:

$$x_i, x_{i+m}, x_{i+2m}, \dots, x_{i+(M+1)m} \quad (9.65)$$

donde M es el mayor entero tal que $i + (M + 1) \cdot m \leq N$.

La sub-secuencia se construye a partir de la secuencia escogiendo los números que se encuentran distanciados entre sí m posiciones, comenzando por el número que ocupa la posición i . El tamaño del “salto”, m , se denomina *desfase*.

Puede calcularse el estimador de la *autocorrelación* de esta sub-secuencia de la manera siguiente³⁰:

$$\hat{\rho}_{i,m} = \frac{1}{M+1} \cdot \left[\sum_{k=0}^M x_{i+k \cdot m} \cdot x_{i+(k+1) \cdot m} \right] - 0.25 \quad (9.66)$$

La *autocorrelación* de los números de una secuencia tiene una interpretación muy intuitiva:

²⁹Los detalles de este test pueden encontrarse, por ejemplo, en el texto (Banks et al. 1996).

³⁰(Banks et al. 1996).

- Si $\rho_{i,m} > 0$, entonces se dice que la sub-secuencia presenta *autocorrelación positiva*. En este caso, los sucesivos valores de la misma tienen una probabilidad mayor de la esperada de tener un valor cercano, es decir, valores grandes son seguidos de valores grandes y valores pequeños son seguidos de valores pequeños.
- Si $\rho_{i,m} < 0$, entonces se dice que la sub-secuencia presenta *autocorrelación negativa*. Esto significa que números grandes tienen tendencia a ir seguidos de números pequeños y viceversa.
- Si $\rho_{i,m} = 0$, entonces se dice que la sub-secuencia *no está autocorrelacionada*, es decir, no presenta ninguno de los dos tipos de dependencia anteriores.

Puesto que un valor de la correlación diferente de cero indica dependencia, el test se plantea de la forma siguiente:

$$\begin{aligned} H_0 : \rho_{i,m} &= 0 \\ H_1 : \rho_{i,m} &\neq 0 \end{aligned}$$

Para valores grandes de M , la distribución del estimador de $\rho_{i,m}$, representado $\hat{\rho}_{i,m}$, está distribuido aproximadamente $N(0, \sigma_{\hat{\rho}_{i,m}}^2)$. Por consiguiente, el estadístico del test es el siguiente:

$$z_0 = \frac{\hat{\rho}_{i,m}}{\sqrt{\sigma_{\hat{\rho}_{i,m}}^2}} \quad (9.67)$$

el cual está distribuido normalmente, con media cero y varianza uno, bajo la suposición de independencia de la secuencia, para M grande.

La desviación estándar del estimador de la autocorrelación puede estimarse de la forma siguiente³¹:

$$\sqrt{\sigma_{\hat{\rho}_{i,m}}^2} = \frac{\sqrt{13 \cdot M + 7}}{12 \cdot (M + 1)} \quad (9.68)$$

La condición del test, para el nivel de significación α deseado, es la siguiente:

$$\begin{aligned} \text{Si: } & -z_{1-\frac{\alpha}{2}} \leq z_0 \leq z_{1-\frac{\alpha}{2}} && \text{entonces el test no refuta } H_0 \\ \text{Si: } & -z_{1-\frac{\alpha}{2}} > z_0 \text{ ó } z_0 > z_{1-\frac{\alpha}{2}} && \text{entonces el test refuta } H_0 \end{aligned} \quad (9.69)$$

A continuación se muestra un ejemplo de aplicación del test.

Ejemplo 9.10.7. Dada la secuencia mostrada a continuación, se pretende contrastar si los números que ocupan las posiciones 3, 8, 13, etc. están correlacionados. Se desea aplicar el test con un nivel de significación igual a 0.05.

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 0.12 | 0.01 | 0.23 | 0.28 | 0.89 | 0.31 | 0.64 | 0.28 | 0.83 | 0.93 |
| 0.99 | 0.15 | 0.33 | 0.35 | 0.91 | 0.41 | 0.60 | 0.27 | 0.75 | 0.88 |
| 0.68 | 0.49 | 0.05 | 0.43 | 0.95 | 0.58 | 0.19 | 0.36 | 0.69 | 0.87 |

En este caso, el comienzo de la sub-secuencia es $i = 3$, el desfase es $m = 5$, el número de elementos de la secuencia completa es $N = 30$, y el mayor entero M tal que $3 + (M + 1) \cdot 5 \leq 30$ es $M = 4$.

El estimador de la autocorrelación y su desviación estándar valen:

$$\begin{aligned} \hat{\rho}_{3,5} &= \frac{1}{4+1} \cdot [0.23 \cdot 0.28 + 0.28 \cdot 0.33 + 0.33 \cdot 0.27 + 0.27 \cdot 0.05 + 0.05 \cdot 0.36] - 0.25 \\ &= -0.1945 \end{aligned} \quad (9.70)$$

$$\sigma_{\hat{\rho}_{3,5}} = \frac{\sqrt{13 \cdot 4 + 7}}{12 \cdot (4 + 1)} = 0.1280 \quad (9.71)$$

³¹(Banks et al. 1996).

El estadístico del test toma el valor:

$$z_0 = \frac{-0.1945}{0.1280} = -1.516 \quad (9.72)$$

Puesto que el correspondiente valor crítico de la distribución normal estándar es $z_{0.975} = 1.960$, este test no rechaza la hipótesis de independencia. \square

Puede observarse que este test no es muy sensible para valores pequeños de M , particularmente cuando los números de la sub-secuencia son pequeños. Por ejemplo, si todos los números de la sub-secuencia del Ejemplo 9.10.7 fueran iguales a cero, se obtendría $\hat{\rho}_{3,5} = -0.25$ y $z_0 = \frac{-0.25}{0.1280} = -1.95$, con lo cual tampoco se rechazaría la hipótesis de independencia.

Pueden formarse muchas sub-secuencias de una determinada secuencia de datos si ésta tiene un valor de N elevado. Por ejemplo, comenzando con el primer elemento de la secuencia, las posibilidades incluyen:

- La secuencia de todos los números.
- La sub-secuencia formada por el primero, tercero, quinto, etc.
- La sub-secuencia formada por el primero, cuarto, etc.
- Etc.

Al realizarse múltiples tests debe tenerse en cuenta que la probabilidad de que alguno de ellos rechace la hipótesis nula por azar, siendo ésta verdadera, aumenta con el número de tests.

Por ejemplo, si se aplica un test con $\alpha = 0.05$ a 10 sub-secuencias independientes, la probabilidad de no encontrar correlación significativa en ninguno de los 10 tests, en caso en que verdaderamente no la haya, es $(0.95)^{10} = 0.60$. Esto significa que existe una probabilidad del 40% de detectar correlación en alguno de los 10 tests cuando en realidad no la hay. Si se escoge $\alpha = 0.10$, y se realizan los 10 tests, la probabilidad de encontrar correlación, únicamente por azar, es del 65%.

La conclusión de ello es que cuando se realizan numerosos tests para contrastar la autocorrelación, ésta puede eventualmente detectarse, únicamente por azar, aunque verdaderamente no exista.

Test de los huecos (gap test)

Este test se emplea para contrastar la significación de los intervalos existentes entre las sucesivas ocurrencias del mismo dígito en una secuencia de dígitos. Considérese, por ejemplo, la siguiente secuencia de dígitos:

4 1 **3** 5 1 7 2 8 2 0 7 9 1 **3** 5 2 7 9 4 1 6 **3**
3 9 6 **3** 4 8 2 **3** 1 9 4 4 6 8 4 1 **3** 8 9 5 5 7
3 9 5 9 8 5 **3** 2 2 **3** 7 4 7 0 **3** 6 **3** 5 9 9 5 5
 5 0 4 6 8 0 4 7 0 **3** **3** 0 9 5 7 9 5 1 6 6 **3** 8
 8 8 9 2 9 1 8 5 4 4 5 0 2 **3** 9 7 1 2 0 **3** 6 **3**

Para facilitar las explicaciones, se ha subrayado el dígito 3. Existen 18 dígitos 3 en la secuencia, los cuales definen 17 “huecos”:

- El primer hueco tiene longitud 10 (número de dígitos entre el primer dígito 3 y el segundo).
- El segundo hueco tiene longitud 7.
- Y así sucesivamente.

Puede calcularse la probabilidad de obtener un hueco de longitud 10 de la forma siguiente:

$$\begin{aligned} \text{Prob}\{\text{Hueco de longitud } 10\} &= \underbrace{\text{Prob}\{\text{no } 3\} \cdot \dots \cdot \text{Prob}\{\text{no } 3\}}_{10 \text{ veces}} \cdot \text{Prob}\{3\} \\ &= (0.9)^{10} \cdot (0.1) \end{aligned} \quad (9.73)$$

Este resultado es generalizable. La probabilidad de obtener un hueco de longitud x es:

$$\text{Prob}\{\text{hueco de longitud } x\} = (0.9)^x \cdot (0.1) \quad \text{para } x : 0, 1, 2, \dots \quad (9.74)$$

En el ejemplo anterior sólo se ha considerado el dígito 3. Sin embargo, para analizar completamente la independencia de la secuencia, se deben analizar todos los dígitos: 0, 1, ..., 9.

Las frecuencias observadas para todos los dígitos son comparadas con las frecuencias teóricas usando el test de Kolmogorov-Smirnov.

La distribución teórica para dígitos ordenados aleatoriamente es:

$$\begin{aligned} \text{Prob}\{\text{hueco de longitud } \leq x\} &= F_X(x) \\ &= 0.1 \cdot \sum_{n:0}^x 0.9^n \\ &= 1 - 0.9^{x+1} \end{aligned} \quad (9.75)$$

El test, aplicado a un dígito, consta de los pasos indicados a continuación. Cuando se aplica el test a números aleatorios, los intervalos

$$\left[0, \frac{1}{10}\right), \left[\frac{1}{10}, \frac{2}{10}\right), \dots, \left[\frac{9}{10}, 1\right) \quad (9.76)$$

juegan el papel de los dígitos aleatorios.

1. Seleccionar las clases del test y calcular la probabilidad acumulada teórica empleando la Ecuación 9.75.
2. Ordenar los huecos observados en la secuencia, para todos los dígitos, en una función de probabilidad acumulada empírica, $F_X^*(x)$, con las mismas clases.
3. Calcular D , como la máxima desviación entre $F_X^*(x)$ y $F_X(x)$.
4. Determinar el valor crítico D_α (de la Tabla T.3, situada al final del libro), para el valor deseado de α y el tamaño de la muestra, N .
5. Si el valor calculado de D es mayor que el valor crítico, D_α , entonces el test rechaza la hipótesis de independencia. En caso contrario, el test no rechaza la hipótesis.

A continuación se muestra un ejemplo de aplicación del test.

Ejemplo 9.10.8. Basándose en la frecuencia con que ocurren los huecos, analizar la independencia de la secuencia de 110 dígitos anterior, con $\alpha = 0.05$.

En la siguiente tabla se muestra la aplicación del test:

| Longitud del hueco | Frecuencia | Frecuencia relativa | Frecuencia relativa acumulada | $F_X(x)$ | $ F_X(x) - F_X^*(x) $ |
|--------------------|------------|---------------------|-------------------------------|----------|-----------------------|
| 0 - 3 | 35 | 0.35 | 0.35 | 0.3439 | 0.0061 |
| 4 - 7 | 22 | 0.22 | 0.57 | 0.5695 | 0.0005 |
| 8 - 11 | 17 | 0.17 | 0.74 | 0.7176 | 0.0224 |
| 12 - 15 | 9 | 0.09 | 0.83 | 0.8147 | 0.0153 |
| 16 - 19 | 5 | 0.05 | 0.88 | 0.8784 | 0.0016 |
| 20 - 23 | 6 | 0.06 | 0.94 | 0.9202 | 0.0198 |
| 24 - 27 | 3 | 0.03 | 0.97 | 0.9497 | 0.0223 |
| 28 - 31 | 0 | 0.00 | 0.97 | 0.9657 | 0.0043 |
| 32 - 35 | 0 | 0.00 | 0.97 | 0.9775 | 0.0075 |
| 36 - 39 | 2 | 0.02 | 0.99 | 0.9852 | 0.0043 |
| 40 - 43 | 0 | 0.00 | 0.99 | 0.9903 | 0.0003 |
| 44 - 47 | 1 | 0.01 | 1.00 | 0.9936 | 0.0064 |

El estadístico vale: $D = \max |F_X(x) - F_X^*(x)| = 0.0224$, y el valor crítico: $D_{0.05} = \frac{1.36}{\sqrt{100}} = 0.136$. Puesto que D es menor que el valor crítico, este test no rechaza la hipótesis de independencia. \square

Test del poker

El *test del poker* se basa en la frecuencia con que ciertos dígitos se repiten en las series de números. El siguiente ejemplo muestra un número de repeticiones anormalmente alto, ya que en todos los casos el número contiene dos dígitos iguales:

0.255 0.577 0.331 0.414 0.828 0.909 0.303 0.001 ...

En números de 3 dígitos, existen únicamente las 3 posibilidades siguientes:

- *Los tres dígitos son diferentes.* La probabilidad de que los 3 dígitos sean diferentes es igual a la probabilidad de que el segundo dígito sea diferente del primero (es decir, 0.9), multiplicada por la probabilidad de que el tercer dígito sea diferente del primero y del segundo (es decir, 0.8):

$$\text{Prob}\{3 \text{ dígitos diferentes}\} = 0.9 \cdot 0.8 = 0.72 \quad (9.77)$$

- *Dos dígitos son iguales.* La probabilidad de que ello suceda es igual a:

$$\text{Prob}\{2 \text{ dígitos iguales}\} = \binom{3}{2} \cdot 0.1 \cdot 0.9 = 0.27 \quad (9.78)$$

- *Los tres dígitos son iguales.* La probabilidad de que se de esta situación es igual a la probabilidad de que el segundo dígito sea igual al primero (es decir, 0.9), multiplicada por la probabilidad de que el segundo dígito sea igual al tercero (es decir, 0.9):

$$\text{Prob}\{3 \text{ dígitos iguales}\} = 0.1 \cdot 0.1 = 0.01 \quad (9.79)$$

El siguiente ejemplo muestra cómo puede aplicarse el test del poker, en conjunción con el test chi-cuadrado, para contrastar la independencia.

Ejemplo 9.10.9. Se ha generado una secuencia de 1000 números de 3 dígitos. Con el fin de aplicar el test del poker, se analizan los números de la secuencia, obteniéndose el siguiente resultado:

- 680 números tienen los 3 dígitos diferentes.
- 289 números contienen 2 dígitos iguales.

- 31 números tienen los 3 dígitos iguales.

En la siguiente tabla se muestra la aplicación del test, con $\alpha = 0.05$:

| Combinación, i | Frecuencia observada, O_i | Frecuencia esperada, E_i | $\frac{(O_i - E_i)^2}{E_i}$ |
|----------------------|-----------------------------|----------------------------|-----------------------------|
| 3 dígitos diferentes | 680 | 720 | 2.22 |
| 3 dígitos iguales | 31 | 10 | 44.10 |
| 2 dígitos iguales | 289 | 270 | 1.33 |
| SUMA | 1000 | 1000 | 47.69 |

El número de grados de libertad es igual al número de clases menos uno. En este caso, 2 grados de libertad.

Puesto que $\chi_{2,0.95}^2 = 5.991 < 47.69$ (ver la Tabla T.2, al final del libro), el test rechaza la hipótesis de que los números son independientes. \square

EJERCICIOS

- 9.1** Aplique el algoritmo de corrección del sesgo, propuesto por von Neumann, a la siguiente secuencia de bits:

0 1 1 0 1 0 0 1 0 1 0 0 1 0 1 1 1 1 0 0 0 1 1 0 1 0 1 0 1

- 9.2** Programe el generador de los cuadrados medios, usando un lenguaje de programación de su elección. A continuación ejecútelo: escoja un número de 4 dígitos como semilla, y genere una secuencia de 10 números pseudo aleatorios.
- 9.3** Programe el generador congruencial lineal, usando un lenguaje de programación de su elección. Ejecute el programa para obtener una secuencia de 10 números pseudo aleatorios a partir del generador siguiente: $a = 17$, $c = 43$ y $m = 100$. Emplee como semilla: $z_0 = 27$.
- 9.4** Realice un programa que, trabajando conjuntamente con el programa que ha realizado en el Problema 9.3, calcule el periodo del generador. Aplíquelo al caso particular siguiente: usando el método congruencial multiplicativo, calcule el periodo del generador definido por los siguientes parámetros: $a = 13$, $m = 2^6 = 64$, para los siguientes 4 valores de la semilla: $z_0 = 1, 2, 3$ y 4 .
- 9.5** Programe el generador de Fibonacci, usando un lenguaje de programación de su elección, y empléelo para generar 100 números pseudo aleatorios con: $m = 1000$, $z_0 = 1$ y $z_1 = 1$. A continuación, aplique los siguientes tests estadísticos, con $\alpha = 0.05$:
- Test chi-cuadrado.
 - Test de Kolmogorov-Smirnov.
 - Ajuste una distribución uniforme a la secuencia generada, empleando la herramienta *Input Analyzer* de Arena.
 - Someta la secuencia a los siguientes dos tests de las carreras: carreras crecientes y decrecientes, y carreras por encima y por debajo de la media.
 - Someta la secuencia al test de autocorrelación.
- 9.6** Programe un generador de Tausworthe, usando un lenguaje de programación de su elección, de modo que (agrupando los bits) el resultado del programa sea una secuencia de dígitos decimales (números enteros comprendidos entre 0 y 9 inclusive).
Empleando el programa realizado, obtenga una secuencia 100 de dígitos decimales, mediante un generador de Tausworthe con $c_2 = c_4 = 1$, y el resto de coeficientes cero.
Finalmente, aplique a esta secuencia el test de los huecos, con $\alpha = 0.05$, para contrastar la hipótesis de independencia.
- 9.7** Explique cómo aplicar la técnica de la *división simulada* al siguiente generador: $z_{i+1} = 7^5 \cdot z_i$.

Tema 10

Generación de observaciones de variables aleatorias

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los métodos fundamentales para la generación de observaciones de variables aleatorias.
- Aplicar algoritmos para generar observaciones de variables aleatorias continuas, discretas y de procesos estocásticos de llegada.

10.1. Introducción

En esta sección se discute la generación de observaciones de una variable aleatoria cuya distribución de probabilidad es conocida. Como se verá a lo largo de la sección, con independencia del algoritmo y de la distribución, en todos los casos, es necesario emplear un generador de números seudo aleatorios.

Dado que existen varios métodos para la generación de observaciones de variables aleatorias, a continuación se indican algunos factores que pueden influir en la decisión de usar uno u otro método en determinado estudio de simulación:

- *Exactitud.* Es deseable que el algoritmo proporcione en la medida de lo posible (dadas las limitaciones de precisión del ordenador, la inexactitud del generador de números aleatorios, etc.) observaciones de la variable aleatoria con exactamente la distribución deseada. En la actualidad existen algoritmos con esta propiedad para la mayoría de las distribuciones comúnmente usadas.
- *Eficiencia.* Es deseable que el algoritmo sea de ejecución rápida, y que no requiera de un gran espacio de memoria.
- *Complejidad.* Es deseable que la comprensión conceptual del algoritmo, y su programación, no resulten extremadamente complejos.
- *Robustez.* Es deseable que el algoritmo sea igualmente eficiente para cualquier valor admisible de los parámetros de la distribución.

En este Tema se discuten los procedimientos más ampliamente usados para la generación de observaciones de variables aleatorias: la *transformación inversa*, las técnicas de *composi-*

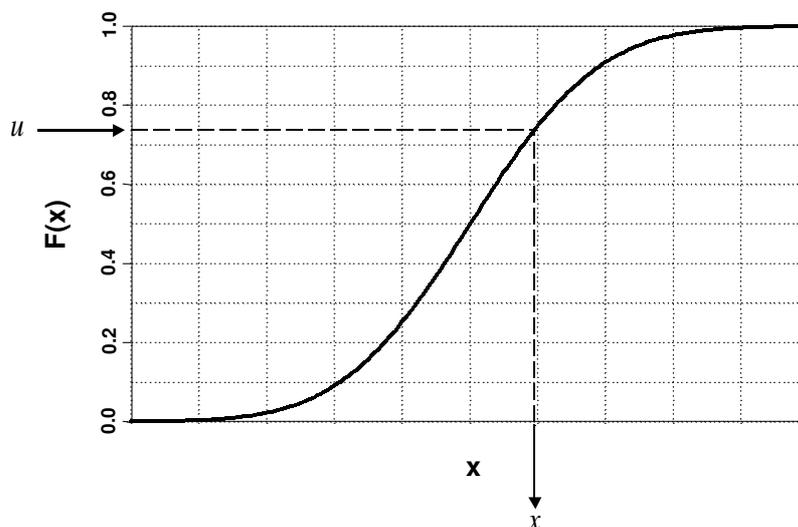


Figura 10.1: Método de la transformación inversa para variables continuas.

ción y de *convolución*, y el método de *aceptación/rechazo*. Asimismo, se discuten algoritmos específicos aplicables a las familias de distribuciones continuas y discretas más comúnmente usadas.

10.2. Método de la transformación inversa

La exposición de los fundamentos del método se ha estructurado en función del tipo de distribución de la variable aleatoria. De esta forma, se han distinguido tres casos: distribuciones continuas, distribuciones continuas truncadas, y distribuciones discretas. A continuación se discute cada uno de ellos.

Aplicación a variables aleatorias continuas

El objetivo es generar observaciones de una variable aleatoria continua, X , cuya probabilidad acumulada, $F_X(x)$, es una función *continua* y *estrictamente creciente* para todos los valores de x tales que $0 < F_X(x) < 1$. El método de la transformación inversa consiste en los dos siguientes pasos (ver la Figura 10.1):

1. Generar un número pseudo aleatorio, u .
2. Devolver $x = F_X^{-1}(u)$, donde F_X^{-1} es la función inversa de F_X .

Obsérvese que, puesto que $F_X(x)$ es una función continua y estrictamente creciente, entonces $F_X^{-1}(u)$ está definido siempre y es único, para valores de u en el intervalo $[0, 1]$.

La demostración del método se basa en que, como $F_X(x)$ es una función monótona creciente, la desigualdad $x_i \leq x_j$ es equivalente a $F_X(x_i) \leq F_X(x_j)$:

$$\text{Prob}\{X \leq x\} = \text{Prob}\{F_X(F_X^{-1}(U)) \leq F_X(x)\} = \text{Prob}\{U \leq F_X(x)\} \quad (10.1)$$

Una dificultad en la aplicación del método a algunas distribuciones continuas, como por ejemplo las distribuciones normal y gamma, es que la inversa de la probabilidad acumulada,

F_X^{-1} no tiene expresión analítica. En estos casos deben emplearse métodos numéricos para calcular $x = F_X^{-1}(u)$.

A continuación se muestra un ejemplo de aplicación del método, para la generación de variables aleatorias distribuidas exponencialmente.

Ejemplo 10.2.1. La probabilidad acumulada de la distribución exponencial es la siguiente:

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \rightarrow F_X^{-1}(u) = -\beta \cdot \ln(1 - u) \quad (10.2)$$

El procedimiento para la generación de muestras de una variable aleatoria exponencial es el siguiente:

1. Generar un número pseudo aleatorio, u .
2. Calcular la observación de la variable aleatoria exponencial:

$$x = -\beta \cdot \ln(u) \quad (10.3)$$

Obsérvese que en la Ecuación 10.3 se usa u , en lugar de $1 - u$, que es lo que aparece en la Ecuación 10.2. La razón de ello es optimizar el cálculo. La justificación es que si la variable aleatoria U está distribuida uniformemente en el intervalo unidad, entonces $1 - U$ también lo está. \square

Aplicación a distribuciones continuas truncadas

El método es fácilmente aplicable a distribuciones truncadas. La densidad de probabilidad truncada al intervalo $[a, b]$ de la variable aleatoria X , y la correspondiente probabilidad acumulada, pueden calcularse a partir de la probabilidad sin truncar de la forma siguiente:

$$f_X^*(x) = \begin{cases} \frac{f_X(x)}{F_X(b) - F_X(a)} & \text{si } a \leq x \leq b \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (10.4)$$

$$F_X^*(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x \end{cases} \quad (10.5)$$

El algoritmo para generar observaciones distribuidas $F_X^*(x)$ es el siguiente:

1. Generar un número pseudo aleatorio, u .
2. Calcular:

$$v = F_X(a) + (F_X(b) - F_X(a)) \cdot u \quad (10.6)$$

Obsérvese que: $V \sim U(F_X(a), F_X(b))$.

3. Devolver:

$$x = F_X^{-1}(v) \quad (10.7)$$

Aplicación a variables aleatorias discretas

Considérese una variable aleatoria discreta X . La probabilidad acumulada de X se define de la forma siguiente:

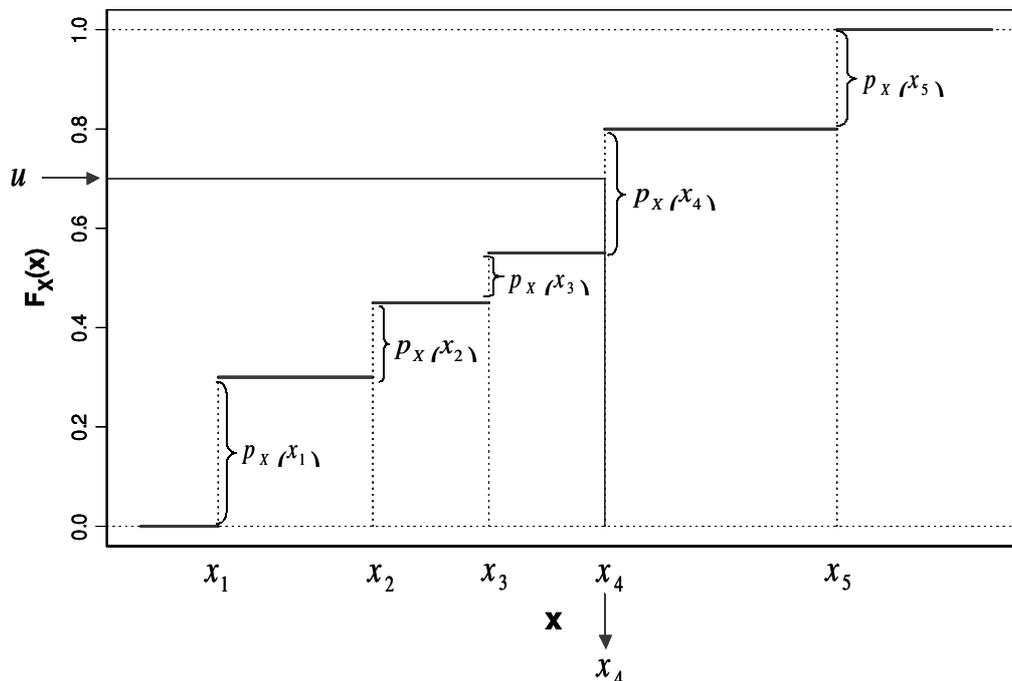


Figura 10.2: Método de la transformación inversa para variables discretas.

$$F_X(x) = \text{Prob}\{X \leq x\} = \sum_{\{i: x_i \leq x\}} p_X(x_i) \quad (10.8)$$

donde: $p_X(x_i) = \text{Prob}\{X = x_i\}$.

El rango (es decir, el conjunto de valores posibles) de la variable aleatoria X es el conjunto de valores $\{x_1, x_2, \dots\}$, los cuales están ordenados crecientemente $x_1 < x_2 < \dots$.

La generación de observaciones de la variable aleatoria X , empleando el método de la transformación inversa, se realiza de la forma siguiente:

1. Generar un número pseudoaleatorio, u .
2. Determinar el menor número entero positivo, I , tal que: $u \leq F_X(x_I)$. Devolver el valor x_I .

Este método es muy intuitivo (ver la Figura 10.2): se divide el intervalo $[0, 1]$ en subintervalos contiguos de longitud $p_X(x_1), p_X(x_2), \dots$. El valor de X se calcula en función del subintervalo en el cual está contenido el número pseudoaleatorio: si u está contenido en el subintervalo i -ésimo (de longitud $p_X(x_i)$), entonces se devuelve el valor x_i .

La demostración del método consiste en comprobar que la probabilidad de obtener x_i aplicando el método es efectivamente $p_X(x_i)$, para todo $\{x_1, x_2, \dots\}$. La comprobación es la siguiente:

- Caso $i = 1$. Se obtiene x_1 si y sólo si

$$u \leq F_X(x_1) = p_X(x_1) \quad (10.9)$$

con lo cual la probabilidad de obtener x_1 es la correcta.

- Caso $i \geq 2$. El algoritmo devuelve x_i si y sólo si

$$F_X(x_{i-1}) < u \leq F_X(x_i) \quad (10.10)$$

por lo que la probabilidad de obtener x_i es:

$$\begin{aligned}\text{Prob}\{X = x_i\} &= \text{Prob}\{F_X(x_{i-1}) < U \leq F_X(x_i)\} \\ &= F_X(x_i) - F_X(x_{i-1}) \\ &= p_X(x_i)\end{aligned}\tag{10.11}$$

A continuación se muestra un ejemplo de aplicación del método.

Ejemplo 10.2.2. En un modelo de la gestión del inventario, el tamaño de los pedidos se representa mediante una variable aleatoria discreta, X , de rango $\{1, 2, 3, 4\}$ y con probabilidad:

$$p_X(1) = \frac{1}{6} \quad p_X(2) = \frac{1}{3} \quad p_X(3) = \frac{1}{3} \quad p_X(4) = \frac{1}{6}\tag{10.12}$$

Para generar realizaciones de esa variable aleatoria discreta, se sigue el procedimiento siguiente:

1. Se genera un número pseudo aleatorio, u .
2. Se devuelve la observación de la variable aleatoria discreta, de acuerdo con el criterio siguiente:

$$u \leq \frac{1}{6} \rightarrow x = 1\tag{10.13}$$

$$\frac{1}{6} < u \leq \frac{1}{2} \rightarrow x = 2\tag{10.14}$$

$$\frac{1}{2} < u \leq \frac{5}{6} \rightarrow x = 3\tag{10.15}$$

$$\frac{5}{6} < u \rightarrow x = 4\tag{10.16}$$

□

10.3. Composición

El método es aplicable cuando la probabilidad acumulada de la variable aleatoria, $F(x)$, puede expresarse como combinación lineal de las probabilidades acumuladas de otras variables aleatorias:

$$F(x) = \sum_{j:1}^{\infty} p_j \cdot F_j(x)\tag{10.17}$$

donde, posiblemente p_j vale cero a partir de cierto valor de j ; y donde $\sum_{j:1}^{\infty} p_j = 1$.

En esta situación, pueden generarse observaciones de la distribución $F(x)$ mediante el siguiente algoritmo, denominado *algoritmo de composición*:

1. Generar un número entero positivo aleatorio, j , tal que: $\text{Prob}\{J = j\} = p_j$. Este número se emplea para seleccionar la distribución j -ésima del conjunto $\{F_1(x), F_2(x), \dots\}$.
2. Generar una observación de la distribución $F_j(x)$. Devolver el número generado.

10.4. Convolución

En algunas distribuciones de probabilidad importantes, la variable aleatoria de la cual se desea obtener observaciones, X , puede expresarse como la suma de determinado número m de otras variables aleatorias independientes:

$$X = Y_1 + Y_2 + \cdots + Y_m \quad (10.18)$$

En estos casos, puede emplearse el *algoritmo de convolución* para obtener observaciones de la variable aleatoria X . El procedimiento es el descrito a continuación:

1. Generar una observación independiente de cada una de las variables aleatorias Y_1, Y_2, \dots, Y_m . Los valores obtenidos se representan: y_1, y_2, \dots, y_m .
2. Devolver $x = y_1 + y_2 + \cdots + y_m$.

10.5. Método de aceptación/rechazo

El siguiente ejemplo pretende ilustrar cuál es el fundamento de la técnica de aceptación/rechazo.

Ejemplo 10.5.1. *Considérese el siguiente procedimiento para generar observaciones de variables aleatorias X distribuidas uniformemente entre $\frac{1}{4}$ y 1:*

1. Generar un número pseudo aleatorio, u .
2. Si $u \geq \frac{1}{4}$, entonces ir al paso 3. En caso contrario, rechazar u y volver al paso 1.
3. Devolver $x = u$. Si se desean obtener más observaciones de X , ir al paso 1, en caso contrario, terminar.

Es decir, cada número pseudo aleatorio generado es “aceptado” o “rechazado”, atendiendo a un determinado criterio: ser mayor o igual que $\frac{1}{4}$ o no serlo. La distribución de los números x devueltos en el paso 2 es $U(0, 1)$ condicionada al evento $\{U \geq \frac{1}{4}\}$. Así pues, la probabilidad de obtener, aplicando este método, números comprendidos entre a y b , donde $\frac{1}{4} \leq a < b \leq 1$ es la siguiente:

$$\text{Prob} \left\{ a < U \leq b \mid \frac{1}{4} \leq U \leq 1 \right\} = \frac{\text{Prob} \{a < U \leq b\}}{\text{Prob} \left\{ \frac{1}{4} \leq U \leq 1 \right\}} = \frac{b - a}{3/4} \quad (10.19)$$

que concuerda con la probabilidad para una distribución uniforme en $[\frac{1}{4}, 1]$. □

Como se ha mostrado en el ejemplo, el concepto básico de la técnica de aceptación y rechazo es el siguiente:

Se generan observaciones de variables aleatorias distribuidas de una determinada forma (en este caso, $U(0, 1)$) hasta que la observación satisface cierta condición (en este caso ser mayor o igual a $\frac{1}{4}$), en cuyo caso se devuelve el valor obtenido.

Para generar observaciones de la distribución $U(\frac{1}{4}, 1)$, resulta más eficiente emplear el método de la transformación inversa que el método de aceptación/rechazo. Sin embargo, existen distribuciones de probabilidad importantes (normal, gamma, beta, etc.) para las cuales la función inversa de la probabilidad acumulada no tiene expresión analítica, con lo cual hay que emplear métodos numéricos para aplicar el método de la transformación inversa. En esta situación, es más eficiente generar observaciones empleando el método de aceptación/rechazo que la transformación inversa. Por ejemplo, en la Sección 10.7 se aplica la técnica de aceptación/rechazo a la generación de variables aleatorias de Poisson.

10.6. Observaciones de variables aleatorias continuas

En esta sección se exponen algunos algoritmos para la generación de observaciones de variables aleatorias continuas¹.

$$X \sim U(a, b)$$

- Generar un número pseudo aleatorio, u .
- Devolver

$$x = a + (b - a) \cdot u \quad (10.20)$$

$$X \sim \text{expo}(\beta)$$

- Generar un número pseudo aleatorio, u .
- Devolver

$$x = -\beta \cdot \ln(u) \quad (10.21)$$

$$X \sim N(\mu, \sigma^2)$$

- Aplicar el *algoritmo de Box y Muller* para generar dos observaciones distribuidas $N(0, 1)$:
 - Generar dos números pseudo aleatorios, u_1 y u_2 .
 - z_1 y z_2 , calculados de la forma siguiente:

$$z_1 = \sqrt{-2 \cdot \ln(u_1)} \cdot \cos(2 \cdot \pi \cdot u_2) \quad (10.22)$$

$$z_2 = \sqrt{-2 \cdot \ln(u_1)} \cdot \sin(2 \cdot \pi \cdot u_2) \quad (10.23)$$

son observaciones independientes de la distribución $N(0, 1)$.

- Cada una de estas dos observaciones, z_1 y z_2 , puede transformarse en una observación de la distribución $N(\mu, \sigma^2)$ de la forma siguiente:

$$x = \mu + \sigma \cdot z \quad (10.24)$$

$$X \sim \text{LN}(\mu, \sigma^2)$$

- Generar una observación, y , de una variable aleatoria, Y , distribuida $N(\mu, \sigma^2)$.
- Devolver

$$x = e^y \quad (10.25)$$

¹En el texto (Law & Kelton 2000) pueden encontrarse algoritmos para las distribuciones gamma, Weibull y beta.

$\text{triang}(a, b, c)$

- Generar un número pseudo aleatorio, u .
- Generar una observación distribuida $\text{triang}(0, 1, c^*)$, con $c^* = \frac{c-a}{b-a}$, mediante la transformación siguiente:

$$y = \begin{cases} \sqrt{c^* \cdot u} & \text{si } u \leq c^* \\ 1 - \sqrt{(1 - c^*) \cdot (1 - u)} & \text{en cualquier otro caso} \end{cases} \quad (10.26)$$

- Generar una observación distribuida $\text{triang}(a, b, c)$ usando la propiedad siguiente:

$$\begin{matrix} Y \sim \text{triang}\left(0, 1, \frac{c-a}{b-a}\right) \\ X = a + (b-a) \cdot Y \end{matrix} \left| \rightarrow X \sim \text{triang}(a, b, c) \quad (10.27)$$

es decir, mediante la transformación siguiente: $x = a + (b-a) \cdot y$

Distribución
empírica
(observaciones
individuales,
 x_1, \dots, x_n)

- Generar un número pseudo aleatorio, u .
- Ordenar crecientemente las observaciones x_1, \dots, x_n . Las observaciones ordenadas se representan $x_{(1)}, \dots, x_{(n)}$, es decir, la observación con número de orden i se representa $x_{(i)}$.
- Devolver:

$$x = x_{(i)} + (p - i + 1) \cdot (x_{(i+1)} - x_{(i)}) \quad (10.28)$$

donde:

$$p = (n - 1) \cdot u \quad (10.29)$$

$$i = \lfloor p \rfloor + 1 \quad (10.30)$$

$\lfloor z \rfloor$ representa el mayor entero que es menor o igual que el número real z .

Distribución
empírica
(observaciones
agrupadas)

Se dispone de las observaciones empíricas agrupadas en k intervalos adyacentes:

$$[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k] \quad (10.31)$$

donde el intervalo j -ésimo contiene n_j observaciones.

- Obtener la probabilidad acumulada empírica, $F_X(x)$, a partir de las observaciones agrupadas.
- Generar un número pseudo aleatorio, u .
- Encontrar el entero no negativo j , tal que se satisfice:

$$F_X(a_j) \leq u < F_X(a_{j+1}) \quad \text{donde } 0 \leq j \leq k-1 \quad (10.32)$$

- Devolver:

$$x = a_j + \frac{u - F_X(a_j)}{F_X(a_{j+1}) - F_X(a_j)} \cdot (a_{j+1} - a_j) \quad (10.33)$$

10.7. Observaciones de variables aleatorias discretas

En esta sección se describen algunos algoritmos que permiten muestrear algunas de las distribuciones discretas más comúnmente usadas en simulación.

$X \sim \text{Bernoulli}(b)$

- Generar un número pseudo aleatorio, u .
- Si $u \leq b$, entonces devolver $x = 1$. En caso contrario, devolver $x = 0$.

$X \sim \text{DU}(i, j)$

- Generar un número pseudo aleatorio, u .
- Devolver:

$$x = i + \lfloor (j - i + 1) \cdot u \rfloor \quad (10.34)$$

$X \sim \text{bin}(t, b)$

- Generar t observaciones independientes, y_1, \dots, y_t , distribuidas

$$Y_i \sim \text{Bernoulli}(b) \quad (10.35)$$

- Devolver:

$$x = y_1 + y_2 + \dots + y_t \quad (10.36)$$

$X \sim \text{geom}(b)$

El algoritmo está basado en la siguiente relación entre las variables aleatorias geométricas y las de Bernoulli:

Si se realiza una secuencia de observaciones independientes y_1, y_2, \dots de variables aleatorias distribuidas $Y_i \sim \text{Bernoulli}(b)$, entonces el número de observaciones anteriores a la primera que vale "1", $x = \min \{i : y_i = 1\} - 1$, es una observación de una variable distribuida $X \sim \text{geom}(b)$.

El algoritmo consta de los pasos siguientes:

1. Asignar $i = 0$.
2. Generar un número aleatorio u , independiente de los anteriormente generados.
3. Si $u \leq b$, entonces devolver $x = i$. En caso contrario, reemplazar i por $i + 1$ y volver al Paso 2.

$X \sim \text{negbin}(t, b)$

- Generar t observaciones independientes, y_1, y_2, \dots, y_t , de variables aleatorias distribuidas: $Y_i \sim \text{geom}(b)$.
- Devolver:

$$x = y_1 + y_2 + \dots + y_t \quad (10.37)$$

$X \sim \text{Poisson}(\lambda)$

El algoritmo está basado en la propiedad siguiente:

Sea y_1, y_2, \dots una secuencia de observaciones de variables aleatorias Y_1, Y_2, \dots independientes e idénticamente distribuidas, y sea una variable aleatoria X cuyas realizaciones se calculan de la forma siguiente:

$$x = \text{máx} \left\{ i : \sum_{j:1}^i y_j \leq 1 \right\} \quad (10.38)$$

es decir, x es igual al mayor número entero, i , tal que $\sum_{j:1}^i y_j \leq 1$. Entonces,

$$Y_i \sim \text{expo} \left(\frac{1}{\lambda} \right) \leftrightarrow X \sim \text{Poisson}(\lambda) \quad (10.39)$$

Las observaciones y_1, y_2, \dots , distribuidas $Y_i \sim \text{expo} \left(\frac{1}{\lambda} \right)$, pueden obtenerse a partir de números seudo aleatorios, u_1, u_2, \dots , aplicando la expresión (10.21): $y_i = \frac{1}{\lambda} \cdot \ln(u_i)$. En consecuencia, la fórmula para generar observaciones distribuidas Poisson (λ) es la siguiente:

$$x = \text{máx} \left\{ i : \sum_{j:1}^i y_j \leq 1 \right\} = \text{máx} \left\{ i : \prod_{j:1}^i u_j \geq e^{-\lambda} \right\} \quad (10.40)$$

El algoritmo (de aceptación/rechazo) es:

1. Asignar $a = e^{-\lambda}$, $b = 1$, $i = 0$.
2. Generar un número aleatorio u , independiente de los generados anteriormente.
3. Reemplazar b por $b \cdot u$.
4. Si el nuevo valor de b satisface $b < a$, entonces devolver $x = i$. En caso contrario, reemplazar i por $i + 1$ y volver al paso 2.

Distribución
discreta
arbitraria

Sea X una variable aleatoria discreta, cuyo rango es $\{0, 1, 2, \dots\}$. La probabilidad de que una observación de X valga $x = i$ es igual a: $\text{Prob}\{X = i\} = p_X(i)$, donde i pertenece al rango de X .

Pueden generarse observaciones de X , aplicando el método de la transformada inversa² de la forma siguiente:

- Generar un número seudo aleatorio, u .
- Devolver $x = k$, donde k es el entero no negativo que satisface:

$$\sum_{j:0}^{k-1} p_X(j) \leq u < \sum_{j:0}^k p_X(j) \quad (10.41)$$

10.8. Generación de procesos de llegada

A continuación se describen algunos algoritmos para generar los instantes de llegada t_1, t_2, \dots en los procesos de Poisson.

Proceso de
Poisson
estacionario

En un proceso de Poisson estacionario de frecuencia $\lambda > 0$, los intervalos de tiempo entre llegadas consecutivas siguen una distribución³ $\text{exp}\left(\frac{1}{\lambda}\right)$. Sacando partido de esta propiedad, puede generarse una secuencia de instantes de llegada de la forma siguiente:

1. Asignar $t_0 = 0$ e $i = 1$.
2. Generar un número seudo aleatorio, u , independiente de los anteriormente generados.

3. Devolver:

$$t_i = t_{i-1} - \frac{1}{\lambda} \cdot \ln(u) \quad (10.42)$$

4. Sustituir i por $i + 1$ y volver al paso 2.

Proceso de
Poisson no
estacionario

Se supone que la frecuencia de llegada es una función del tiempo, $\lambda(t)$, y que además está acotada (es decir, existe λ^* tal que $\lambda^* = \max_t \{\lambda(t)\}$, y es finito).

El algoritmo consiste esencialmente en:

- Generar un proceso de Poisson estacionario, con frecuencia λ^* . Sea $\{t_1^*, t_2^*, \dots\}$ la secuencia de instantes de llegada obtenidos.
- A continuación, se eliminan instantes de llegada de la secuencia $\{t_1^*, t_2^*, \dots\}$, atendiendo al criterio aleatorio siguiente:

La probabilidad de eliminar el instante t_i^* de la secuencia es:

$$1 - \frac{\lambda(t_i^*)}{\lambda^*} \quad (10.43)$$

De esta forma, es más probable eliminar el instante de llegada t_i^* cuando la frecuencia en ese instante, $\lambda(t_i^*)$, es pequeña. Por el contrario, a medida que la frecuencia en ese instante se aproxima al máximo, λ^* , la probabilidad de eliminar el instante tiende a cero.

El siguiente algoritmo aplica el método anteriormente expuesto:

1. Asignar $t = t_{i-1}$.
2. Generar dos números seudo aleatorios, u_1 y u_2 , independientes entre sí e independientes de los previamente generados.
3. Reemplazar t por:

$$t - \frac{1}{\lambda^*} \cdot \ln(u_1) \quad (10.44)$$

4. Si $u_2 \leq \frac{\lambda(t)}{\lambda^*}$, entonces devolver $t_i = t$. En caso contrario, volver al paso 2.

Llegadas
en grupo

Se supone que las entidades llegan en grupos de dimensión aleatoria. El número de entidades del grupo que llega en el instante t_i es una variable aleatoria discreta B_i .

El algoritmo para generar el tamaño de los grupos se aplica en conjunción con el algoritmo para generar los instantes de llegada:

1. Generar el siguiente instante de llegada, t_i .
2. Generar una realización independiente de la variable aleatoria discreta B_i . El número entero obtenido es el número de entidades que llegan en el instante t_i .

EJERCICIOS

- 10.1** Realice un programa que genere observaciones de una distribución exponencial a partir de números seudo aleatorios. Emplee el lenguaje de programación que desee.

A continuación, ejecute el programa de generación de números seudo aleatorios para los datos del generador descrito en el Problema 9.3, con la semilla $z_0 = 20$, hasta obtener un periodo completo, y a partir de esa secuencia de números seudo aleatorios, obtenga observaciones de una variable aleatoria distribuida exponencialmente con media 1 minuto, usando el programa que acaba de realizar.

Finalmente, dibuje el histograma de las observaciones exponenciales. Además ajuste una distribución uniforme a la secuencia de números seudo aleatorios, y una distribución exponencial a la secuencia de observaciones exponenciales. Emplee para ello la herramienta *Input Analyzer*.

- 10.2** Se han realizado 100 observaciones del tiempo necesario para reparar una máquina. Los datos obtenidos son los siguientes:

| Tiempo (horas) | Número de observaciones |
|---------------------|-------------------------|
| $0 \leq x \leq 0.5$ | 31 |
| $0.5 < x \leq 1.0$ | 10 |
| $1.0 < x \leq 1.5$ | 25 |
| $1.5 < x \leq 2.0$ | 34 |

- a) A partir de estos datos, defina la probabilidad acumulada empírica.
 - b) Se sabe que, por muy rápido que se realice una reparación, dura al menos 15 minutos. Defina la probabilidad acumulada empírica, truncando el rango de la variable aleatoria al intervalo $[0.5, 2]$ horas.
 - c) Explique detalladamente cómo aplicar el método de la transformación inversa para generar observaciones de esta distribución empírica truncada.
- 10.3** Se ha observado que una variable aleatoria discreta puede tomar los valores 0, 1 y 2, con probabilidad 0.50, 0.20 y 0.30 respectivamente. Describa cómo aplicaría el método de la transformación inversa para generar observaciones de esa distribución empírica discreta.
- 10.4** La probabilidad de una variable aleatoria discreta X es la siguiente:

$$p_X(x) = \frac{2 \cdot x}{k \cdot (k + 1)}, \text{ para } x = 1, 2, \dots, k$$

- a) Calcule la probabilidad acumulada de X .
 - b) Explique detalladamente cómo pueden obtenerse observaciones X mediante el método de la transformación inversa.
- 10.5** Demuestre que el algoritmo de Box y Muller efectivamente genera observaciones de la distribución normal estándar.

10.6 Puede demostrarse que una variable aleatoria X , distribuida Erlang con parámetros (K, θ) , es la suma de K variables aleatorias exponenciales independientes X_i (con $i = 1, \dots, K$), cada una de las cuales tiene una media igual a: $\frac{1}{K \cdot \theta}$. Explique cómo generar observaciones de X mediante el método de la convolución.

10.7 Genere tres observaciones de una distribución de Poisson con media $\lambda = 0.2$. Emplee la siguiente secuencia de números pseudo aleatorios:

0.4357 0.4146 0.8353 0.9952 0.8004

Parte IV

Empleo de los modelos de simulación

Tema 11

Análisis estadístico de los resultados de la simulación

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir y aplicar las técnicas para el análisis de las salidas de un único sistema.
- Discutir y aplicar las técnicas para comparar dos o más sistemas.
- Analizar las salidas de un único sistema usando Arena.
- Comparar las salidas de varios sistemas usando Arena.

11.1. Introducción

Un caso típico de mala planificación de un estudio de simulación es aquel en el cual se invierte una gran cantidad de tiempo y dinero en el desarrollo y programación del modelo, y sin embargo se dedica muy poco esfuerzo a analizar adecuadamente los datos de salida de la simulación: se realiza una única réplica de la simulación, de longitud arbitraria, y se considera (erróneamente) que se han obtenido las características “verdaderas” del modelo.

En realidad, debido al carácter estocástico de la simulación, las estimaciones de las características del modelo obtenidas de la simulación son realizaciones (es decir, observaciones) de variables aleatorias, cuyas distribuciones posiblemente tengan grandes varianzas. En consecuencia, las estimaciones obtenidas en una única réplica de la simulación pueden diferir enormemente de las características “verdaderas” del modelo y, por tanto, conducir a inferencias erróneas acerca del sistema bajo estudio. A continuación se muestra un ejemplo que pretende ilustrar este hecho.

Ejemplo 11.1.1. *Para modelar una oficina de atención al público, en la que trabaja un único empleado, se realizan las hipótesis siguientes:*

$$\text{Intervalo de tiempo entre llegadas sucesivas} \sim \text{expo}\left(\frac{1}{\lambda}\right) \quad (11.1)$$

$$\text{Tiempo de servicio} \sim \text{expo}\left(\frac{1}{\omega}\right) \quad (11.2)$$

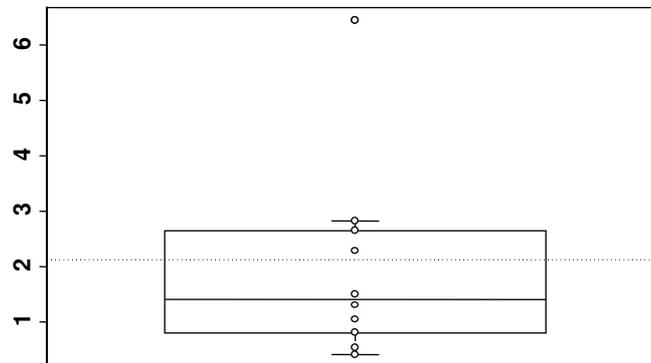


Figura 11.1: Muestras del tiempo medio en cola, $X (m = 25)$.

El objetivo del estudio de simulación es extraer ciertas conclusiones acerca del tiempo medio de espera en la cola de los clientes. Para ello, debe analizarse estadísticamente el proceso estocástico $\{D_i, i \geq 1\}$, donde D_i es el tiempo de espera en la cola del cliente i .

La condición de finalización de la simulación es que m clientes comiencen a ser atendidos. La condición inicial, I , de la simulación es la siguiente: el empleado se encuentra libre, y el número de clientes que está esperando en la cola es $L(0)$.

El objetivo de la simulación es estudiar la distribución de probabilidad de la variable aleatoria "tiempo medio en cola de $m = 25$ clientes", $X(m) = \frac{\sum_{i=1}^m D_i}{m}$. La condición inicial de la simulación es: $L(0) = 0$, y el valor de los parámetros del sistema: $\frac{\lambda}{\omega} = 0.9$.

Las observaciones de la variable aleatoria tiempo medio de espera en cola, $X(m = 25)$, obtenidas en 10 réplicas independientes de la simulación, son las siguientes:

| | | | | |
|-------|-------|-------|-------|-------|
| 1.051 | 6.438 | 2.646 | 0.805 | 1.505 |
| 0.546 | 2.281 | 2.822 | 0.414 | 1.307 |

La media, la varianza y la mediana de las observaciones son:

Media: 1.9815

Varianza: 3.1720

Mediana: 1.4060

En la Figura 11.1 se representa el boxplot de las 10 observaciones del tiempo medio en cola, obtenidas mediante simulación. Los puntos superpuestos al boxplot indican el valor numérico de cada una de las observaciones. La línea horizontal punteada representa el valor esperado exacto del tiempo medio en cola, calculado empleando técnicas analíticas:

$$E\{X(m = 25) | L(0) = 0\} = 2.124$$

Puede observarse que algunas de las observaciones del tiempo medio en cola difieren enormemente del valor esperado "verdadero". La conclusión de ello es que, para poder extraer conclusiones válidas del resultado de la simulación, es necesario aplicar técnicas estadísticas (por ejemplo, el uso de intervalos de confianza para el valor esperado). \square

La simulación es un experimento de muestreo realizado mediante el uso del ordenador. Si se desea extraer resultados válidos de un estudio de simulación, es preciso emplear técnicas estadísticas adecuadas para diseñar y analizar los experimentos de simulación.

Las técnicas a emplear en el diseño y en el análisis del experimento de simulación dependen del tipo de experimento. A su vez, el tipo de experimento depende del objetivo del estudio de simulación.

Los experimentos de simulación pueden clasificarse en los siguientes dos tipos: *con condición de finalización* y *sin condición de finalización*. La diferencia entre ellos radica en que exista o no un evento, E , que establezca de manera “natural” la condición de finalización de cada réplica de la simulación.

- *Simulaciones con condición de finalización*. En las simulaciones con condición de finalización las medidas del comportamiento del modelo son relativas al intervalo de tiempo simulado, $[0, T_E]$, donde T_E es el instante (posiblemente aleatorio) en que sucede el evento de finalización, E . En general, las medidas del comportamiento del modelo durante el intervalo $[0, T_E]$ dependen del estado del modelo en el instante inicial de la simulación.
- *Simulaciones sin condición de finalización*. La finalidad de este tipo de simulaciones es estudiar las características del modelo cuando el tiempo simulado tiende a infinito. En esta situación, las características del modelo son independientes de su estado inicial.

Comportamiento transitorio y estacionario

En las dos definiciones anteriores están implícitos los conceptos de *comportamiento transitorio* y de *comportamiento estacionario* del proceso estocástico de salida:

- En las simulaciones con finalización, se pretenden estudiar determinadas propiedades del proceso estocástico, teniendo en cuenta su comportamiento transitorio.
- Por el contrario, en las simulaciones sin condición de finalización se pretende estudiar el comportamiento estacionario del proceso.

Con el fin de clarificar el significado de los conceptos “*transitorio*” y “*estacionario*”, consideremos un proceso estocástico: Y_1, Y_2, \dots . La probabilidad acumulada de la variable aleatoria Y_i del proceso, condicionada a que el estado inicial del sistema sea I , se representa de la manera siguiente:

$$F_{Y_i}(y_i | I) = \text{Prob} \{Y_i \leq y_i | I\} \quad (11.3)$$

donde $F_{Y_i}(y_i | I)$ es la *distribución transitoria* del proceso estocástico para el valor i del índice temporal y para una condición inicial I del sistema.

En algunos modelos, al tender el tiempo simulado a infinito (es decir, $i \rightarrow \infty$), las distribuciones de las variables aleatorias Y_i tienden a determinada distribución, F , con independencia del estado inicial I del sistema:

$$F_{Y_i}(y | I) \rightarrow F(y) \quad \text{cuando } i \rightarrow \infty \quad (11.4)$$

para cualquier condición inicial, I . Entonces, se dice que F es la *distribución en el estacionario* del proceso estocástico Y_1, Y_2, \dots .

Si bien la distribución en el estacionario se obtiene en el límite $i \rightarrow \infty$, a efectos prácticos puede considerarse que a partir de cierto valor $k + 1$ del índice temporal, la distribución es “esencialmente” la misma. Se dice entonces que el estado estacionario del proceso comienza en el instante $k + 1$.

Obsérvese que el estado estacionario no significa que las observaciones de las variables aleatorias Y_{k+1}, Y_{k+2}, \dots vayan a tomar el mismo valor en una determinada réplica de la simulación. Significa que todas ellas están aproximadamente distribuidas de la misma forma.

La distribución en el estacionario F no depende del estado inicial del sistema, I , sin embargo, la velocidad con que las distribuciones $F_{Y_i}(y_i | I)$ convergen a F normalmente sí depende del estado inicial. A continuación se muestra un ejemplo.

Ejemplo 11.1.2. Se han realizado 100 réplicas independientes de longitud $m = 31$ de la simulación de cierto modelo, todas ellas partiendo del mismo estado inicial del modelo. Sea $y_{1,1}, y_{1,2}, \dots, y_{1,31}$ la realización de las variables aleatorias Y_1, Y_2, \dots, Y_{31} obtenida de la primera réplica de la simulación. Si se realiza una segunda réplica de la simulación, independiente¹ de la anterior, entonces se obtiene una realización diferente $y_{2,1}, y_{2,2}, \dots, y_{2,31}$ de las variables aleatorias Y_1, Y_2, \dots, Y_{31} . Las observaciones obtenidas al realizar las 100 réplicas de la simulación son:

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,31} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,31} \\ \vdots & \vdots & \ddots & \vdots \\ y_{100,1} & y_{100,2} & \cdots & y_{100,31} \end{array} \quad (11.5)$$

En general, las observaciones de una réplica en particular (fila) no son independientes entre sí y además no están idénticamente distribuidas (ya que son realizaciones de diferentes variables aleatorias).

Sin embargo, las observaciones de una misma variable aleatoria (columna), $y_{1,i}, y_{2,i}, \dots, y_{100,i}$, son independientes y están idénticamente distribuidas. La independencia entre las réplicas de la simulación es la clave para poder aplicar los métodos estadísticos (relativamente simples) descritos en este tema. El objetivo del análisis estadístico de este proceso estocástico es usar las observaciones obtenidas mediante simulación para realizar inferencias acerca de las distribuciones de probabilidad de las variables aleatorias Y_1, Y_2, \dots, Y_{31} .

En la Figura 11.2 se ha representado el boxplot de las 100 observaciones de cada una de las variables aleatorias. Se observa que a partir de $k + 1 \approx 22$ el proceso alcanza aproximadamente el estado estacionario. Se puede emplear el test de Kruskal-Wallis para contrastar la hipótesis de que las variables aleatorias $Y_{22}, Y_{23}, \dots, Y_{31}$ están idénticamente distribuidas.

Si se hubieran realizado las réplicas de la simulación partiendo de un estado inicial I diferente, en general las distribuciones transitorias hubieran sido diferentes, así como el valor del índice temporal a partir del cual se hubiera podido considerar que se ha alcanzado (aproximadamente) el estacionario. \square

Existen procesos estocásticos Y_1, Y_2, \dots en los cuales, cuando el tiempo simulado tiende a infinito, las distribuciones de las variables aleatorias no convergen a una determinada distribución en el estado estacionario. A continuación se muestra un ejemplo.

Ejemplo 11.1.3. En la Figura 11.3 se muestran 100 observaciones de cierto proceso estocástico, en el que Y_i representa el tiempo en cola del cliente i , en un servicio de atención al público mal dimensionado: el número de clientes en la cola y el tiempo de espera de éstos aumenta más y más al avanzar el tiempo simulado. \square

También puede suceder que el proceso estocástico Y_1, Y_2, \dots tenga un comportamiento oscilante. Supóngase que se divide el eje temporal en intervalos de igual longitud, denominados “ciclos”, y que se denomina $Y_i^{C_j}$ a la variable aleatoria i del ciclo j . El proceso tiene un comportamiento oscilante si satisface:

- Las distribuciones de las variables $Y_1^{C_j}$ tienden, cuando el número de ciclos tiende a infinito, $j \rightarrow \infty$, a cierta distribución F_1 .
- Las distribuciones de las variables $Y_2^{C_j}$ tienden, cuando el número de ciclos tiende a infinito, $j \rightarrow \infty$, a cierta distribución F_2 .
- Etc.

A efectos prácticos, a partir de determinado ciclo, $k + 1$, puede considerarse que:

- Las variables aleatorias $Y_1^{C_j}$, con $j : k + 1, k + 2, \dots$, están idénticamente distribuidas.

¹Para cada réplica se usa un conjunto de números aleatorios independiente de los conjuntos de números aleatorios usados en las demás réplicas. Además, los contadores estadísticos son inicializados al comienzo de cada réplica. En todas las réplicas se usan las mismas condiciones iniciales.

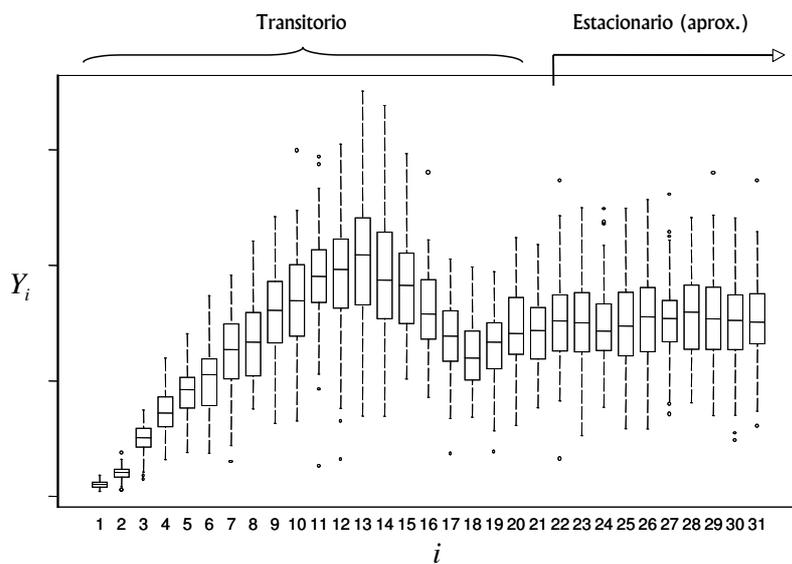


Figura 11.2: Proceso estocástico con estado estacionario.

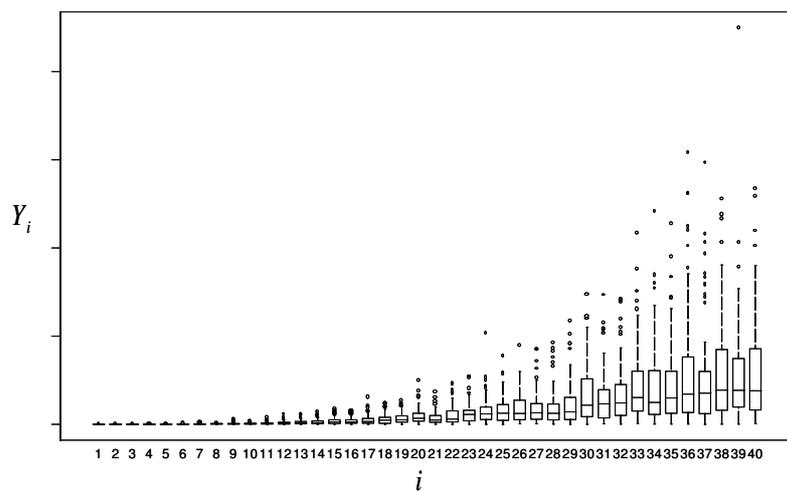


Figura 11.3: Proceso estocástico sin estado estacionario.

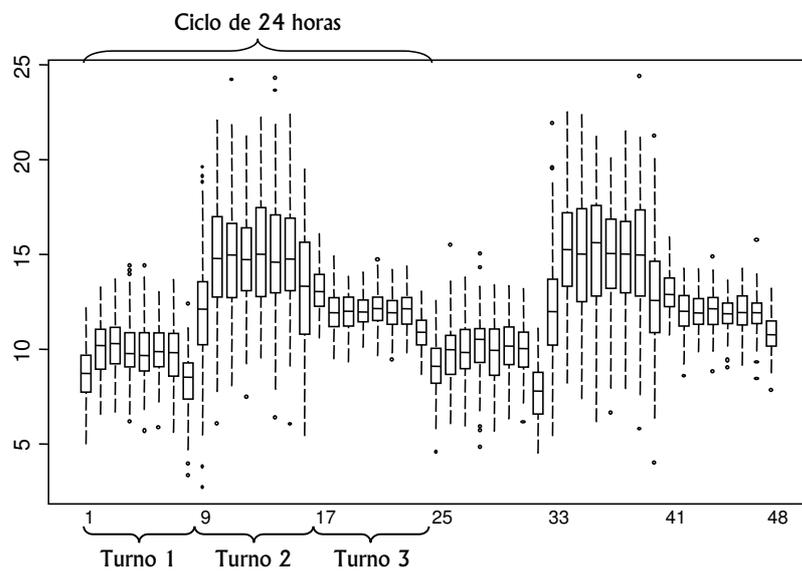


Figura 11.4: Proceso estocástico oscilante.

- Las variables aleatorias $Y_2^{C_j}$, con $j : k + 1, k + 2, \dots$, están idénticamente distribuidas.
- Etc.

A continuación se muestra un ejemplo.

Ejemplo 11.1.4. *Supóngase, por ejemplo, el modelo de una cadena de montaje de electrodomésticos que funciona 24 horas al día. Los trabajadores están distribuidos en 3 turnos de 8 horas cada día. Dado que las condiciones que rodean el trabajo de cada uno de los turnos son diferentes, es de prever que el número de electrodomésticos producidos por hora oscile con un periodo de 24 horas. En la Figura 11.4 se representan 100 muestras del proceso estocástico. Se ha escogido una ventana de tiempo en la cual el sistema ha alcanzado un comportamiento cíclico estacionario.* \square

11.2. Simulaciones con condición de finalización

Considérese de nuevo el estudio de simulación del Ejemplo 11.1.1. En este caso la condición de finalización es diferente: que el tiempo simulado alcance 8 horas. El número total de clientes que abandonan la cola durante la simulación es una variable aleatoria, M . Por tanto, el tiempo medio en cola se define de la forma siguiente:

$$X = \frac{1}{M} \cdot \sum_{i:1}^M D_i \quad (11.6)$$

Para estimar el valor esperado de la variable aleatoria X , μ_X , se realizan n réplicas independientes de la simulación, todas ellas con la misma condición inicial, obteniéndose las n observaciones siguientes de X : x_1, x_2, \dots, x_n . La media aritmética de estas n observaciones, $\bar{X}(n) = \frac{1}{n} \cdot \sum_{i:1}^n x_i$, es un estimador no sesgado de μ_X .

Si la variable aleatoria X está distribuida de forma normal, $X \sim N(\mu_X, \sigma_X^2)$, entonces:

$$\bar{X}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \quad (11.7)$$

es un intervalo del $100 \cdot (1 - \alpha) \%$ de confianza para μ_X . Como se observa en la Ecuación (11.7), la anchura del intervalo de confianza depende, entre otros factores, del tamaño de la muestra, n , que en este caso corresponde con el número de réplicas de la simulación.

Cabe plantearse el problema inverso: ¿cuántas réplicas deberán realizarse para obtener un intervalo de confianza de una determinada anchura? Supóngase que se realizan inicialmente n_0 réplicas de la simulación², obteniéndose n_0 observaciones independientes de X : x_0, \dots, x_{n_0} . A partir de estas observaciones se calcula el intervalo de confianza para la media de X , obteniéndose un intervalo demasiado ancho. ¿Cuántas réplicas de la simulación, n , haría falta realizar para que la anchura del intervalo fuera inferior a un determinado valor, h ?

Puede calcularse un valor aproximado para el número de réplicas de la forma siguiente³:

$$n_1 \cong z_{1-\frac{\alpha}{2}}^2 \cdot \frac{S_X^2(n_0)}{h^2} \quad (11.8)$$

²En (Law & Kelton 2000) se ofrece la siguiente recomendación: al menos deben realizarse entre 3 y 5 réplicas de la simulación, para estimar la variabilidad de las observaciones de X . Si por razones de tiempo o de coste esto no es posible, entonces es preferible no realizar el estudio de simulación, ya que no habrá forma de estimar el error cometido en los resultados.

³Referencia: (Kelton et al. 2002). Obsérvese que el número de réplicas es inversamente proporcional al cuadrado de la anchura del intervalo: para reducir el intervalo a la mitad es necesario disponer de 4 veces más datos. Esto resulta lógico: al crecer el número de réplicas, cada réplica tiene una contribución menor al "conocimiento acumulado".

donde $S_X^2(n_0)$ es el estimador de la varianza de X , calculado de las n_0 observaciones x_0, \dots, x_{n_0} . Obsérvese que para obtener esta aproximación se ha sustituido el valor crítico de la distribución t por el valor crítico de la distribución normal. Ambos valores críticos están próximos para tamaños de muestra superiores a 30.

Una vez realizadas las réplicas adicionales, si el intervalo de confianza obtenido de las n_1 observaciones es demasiado amplio, entonces debe repetirse el cálculo del número de réplicas:

$$n_2 \cong z_{1-\frac{\alpha}{2}}^2 \cdot \frac{S_X^2(n_1)}{h^2} \quad (11.9)$$

donde el estimador de la varianza se calcula de las n_1 observaciones disponibles. Y así sucesivamente.

11.3. Simulaciones en el estacionario

Supóngase que se desea estimar la media de la distribución en el estacionario de un determinado proceso estocástico, Y_1, Y_2, \dots, Y_m . Para realizar este cálculo es preciso excluir la parte transitoria del proceso. Es decir, si se considera que el proceso alcanza el estacionario a partir del índice temporal $k+1$, la media de la distribución en el estacionario se estimará de la forma siguiente:

$$\bar{Y}(m, k) = \frac{1}{m-k} \cdot \sum_{i:k+1}^m y_i \quad (11.10)$$

Surge inmediatamente la cuestión siguiente: ¿cómo deben escogerse los valores de k y m para que el estimador sea no sesgado? El método gráfico de Welch proporciona una respuesta a esta pregunta.

Método gráfico de Welch

El *método de Welch* es una técnica gráfica cuyo propósito es determinar el valor del índice temporal, k , a partir del cual puede considerarse que la media de las variables aleatorias Y_i ha alcanzado aproximadamente un valor estacionario, μ_Y . Es decir:

$$E\{Y_i\} \approx \mu_Y \quad \text{para todo } i > k \quad (11.11)$$

Dada la inherente variabilidad de las observaciones del proceso Y_1, Y_2, \dots, Y_m , el método requiere la realización de n réplicas independientes de la simulación. La técnica de Welch consta de los 4 pasos siguientes:

1. Realizar n réplicas independientes de la simulación (con $n \geq 5$), cada una de longitud m (m debe ser grande). Se obtienen las siguientes observaciones:

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,m} \end{array} \quad (11.12)$$

donde $y_{j,i}$ es la observación de la variable aleatoria Y_i obtenida en la réplica j .

2. A partir de las n observaciones aleatorias de cada variable Y_i , calcular el estimador del valor esperado de Y_i . El cálculo se realiza de la forma siguiente:

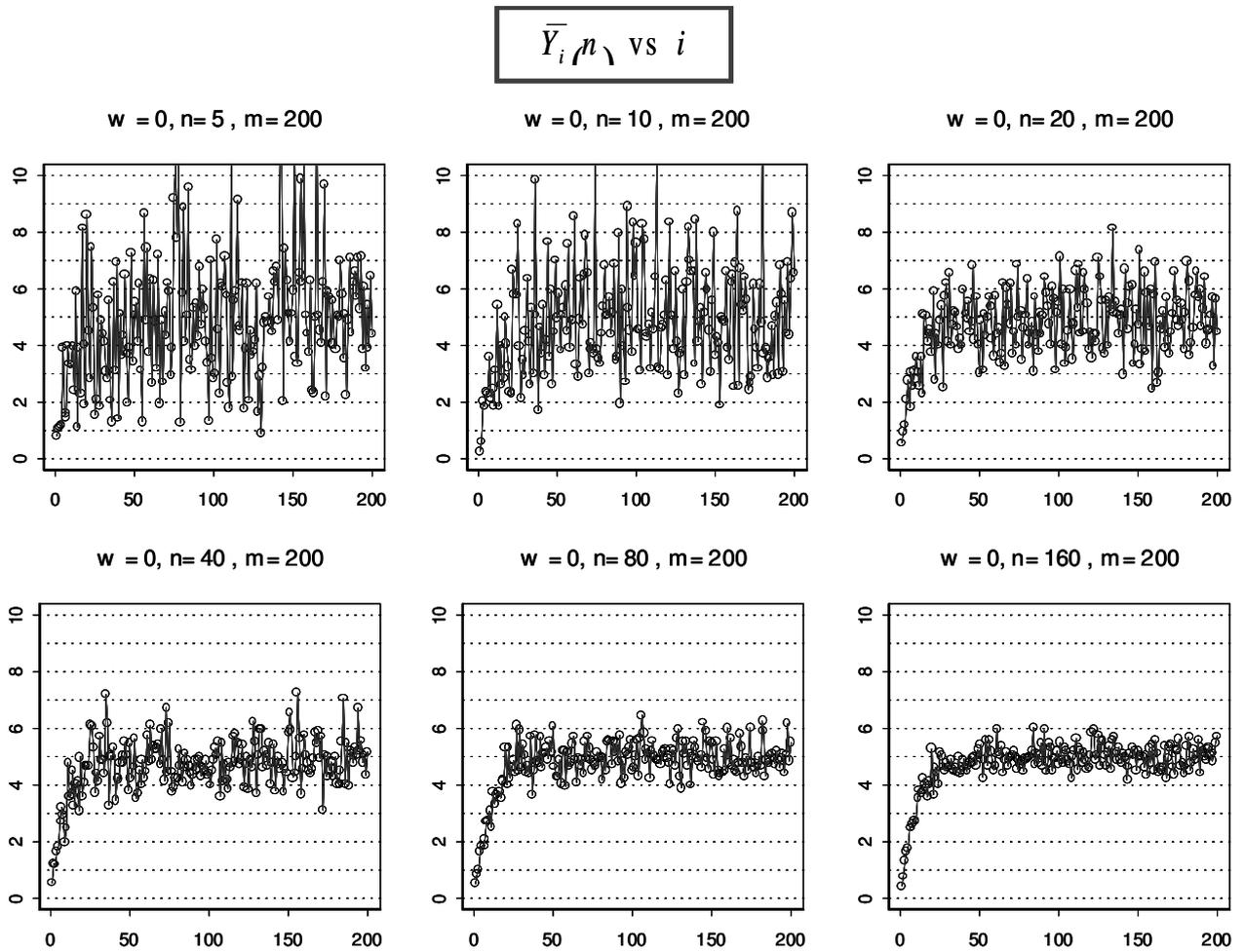


Figura 11.5: Aumentar la muestra disminuye la dispersión del estimador.

$$\bar{Y}_i(n) = \frac{1}{n} \cdot \sum_{j:1}^n y_{j,i} \quad \text{para } i : 1, \dots, m \quad (11.13)$$

La media y la varianza del estimador de la media, $\bar{Y}_i(n)$, son respectivamente:

$$E\{\bar{Y}_i(n)\} = E\{Y_i\} \quad (11.14)$$

$$\text{var}\{\bar{Y}_i(n)\} = \frac{\text{var}\{Y_i\}}{n} \quad (11.15)$$

Cuanto mayor es el número de réplicas de la simulación, n , menor es la varianza del estimador, y por tanto menor es la dispersión de sus observaciones respecto al valor “verdadero”: $E\{Y_i\}$. El siguiente ejemplo ilustra este hecho.

Ejemplo 11.3.1. En la Figura 11.5, aparecen graficados los valores de $\bar{Y}_i(n)$ obtenidos en una simulación de duración $m = 200$, para diferentes valores del número de réplicas, n . Como puede observarse, al aumentar n la dispersión del estimador de la media se reduce. \square

3. Calcular la media móvil de $\bar{Y}_i(n)$, con una ventana w , para cada instante i . La finalidad de ello es suavizar las oscilaciones de “alta frecuencia” en la gráfica $\bar{Y}_i(n)$ vs i , de modo que sea más fácil interpretar la tendencia de “baja frecuencia”.

La media móvil, $\bar{Y}_i(n)|_w$, se calcula de la forma siguiente:

$$\bar{Y}_i(n)|_w = \begin{cases} \frac{\sum_{s:i-w}^w \bar{Y}_{i+s}(n)}{2 \cdot w + 1} & \text{si } i : w + 1, \dots, m - w \\ \frac{\sum_{s:-(i-1)}^{i-1} \bar{Y}_{i+s}(n)}{2 \cdot i - 1} & \text{si } i : 1, \dots, w \end{cases} \quad \text{para } w \leq \left\lfloor \frac{m}{2} \right\rfloor \quad (11.16)$$

Obsérvese que si i no está próximo al comienzo del proceso, entonces $\bar{Y}_i(n)|_w$ es simplemente el promedio de las $2 \cdot w + 1$ observaciones siguientes:

$$\bar{Y}_{i-w}(n), \dots, \bar{Y}_i(n), \dots, \bar{Y}_{i+w}(n) \quad (11.17)$$

Ejemplo 11.3.2. A continuación se indican las expresiones para el cálculo de la media móvil en el caso $w = 2$.

| | | | | | | |
|-------------------|----------------|---------------------------------|---------------------------------|-----|---------------------------------------|-----|
| i | 1 | 2 | 3 | ... | i | ... |
| $Y_i(n)$ | $Y_1(n)$ | $Y_2(n)$ | $Y_3(n)$ | ... | $Y_i(n)$ | ... |
| $\bar{Y}_i(n) _w$ | $\bar{Y}_1(n)$ | $\frac{\sum_{j:1}^3 Y_j(n)}{3}$ | $\frac{\sum_{j:1}^5 Y_j(n)}{5}$ | ... | $\frac{\sum_{j:i-2}^{i+2} Y_j(n)}{5}$ | ... |

| | | | | |
|-------------------|---------------------------------------|-----------------------------------|--------------|----------|
| i | $m - 3$ | $m - 2$ | $m - 1$ | m |
| $Y_i(n)$ | $Y_{m-3}(n)$ | $Y_{m-2}(n)$ | $Y_{m-1}(n)$ | $Y_m(n)$ |
| $\bar{Y}_i(n) _w$ | $\frac{\sum_{j:m-5}^{m-1} Y_j(n)}{5}$ | $\frac{\sum_{j:m-4}^m Y_j(n)}{5}$ | - | - |

□

4. Representar $\bar{Y}_i(n)|_w$ vs i , para $i : 1, 2, \dots, m - w$. Con ayuda de esta gráfica, escoger el valor de k a partir del cual $\bar{Y}_i(n)|_w$ parece alcanzar un nivel constante. A continuación se muestra un ejemplo.

Ejemplo 11.3.3. En la Figura 11.6 se muestran las gráficas $\bar{Y}_i(n)|_w$ vs i del mismo proceso representado en la Figura 11.5. Se ha calculado la media móvil empleando 6 valores diferentes del tamaño de ventana. Puede observarse claramente que el “suavizado” de la curva aumenta con el tamaño de la ventana y facilita la localización del instante a partir del cual comienza el comportamiento estacionario. □

Pueden ofrecerse⁴ las siguientes recomendaciones para la elección de los parámetros n , m y w :

- Realizar inicialmente $n = 5$ ó $n = 10$ réplicas de la simulación (en función de la carga computacional que tenga la ejecución del modelo), tan largas (es decir, con m tan grande) como sea razonablemente posible. En particular, m debe ser mucho más grande que el valor que se presuponga que tiene k (índice temporal donde comienza aproximadamente el estado estacionario), y además m debe ser lo suficientemente grande como para permitir que eventos infrecuentes (por ejemplo, la rotura de maquinas, etc.) ocurran un número razonable de veces.
- Representar $\bar{Y}_i(n)|_w$ vs i para varios valores del tamaño de ventana, w . Escoger el valor de w más pequeño para el cual la gráfica es “razonablemente suave”. Usar esa gráfica para estimar (visualmente) el valor de k .
- Si ninguno de los valores de w del Paso 3 del método de Welch es satisfactorio, realizar 5 ó 10 réplicas adicionales de longitud m . Repetir el Paso 2 empleando todas las observaciones disponibles. Para un valor fijo de w , el gráfico $\bar{Y}_i(n)|_w$ vs i será más suave al aumentar el número de réplicas, n .

⁴(Law & Kelton 2000).

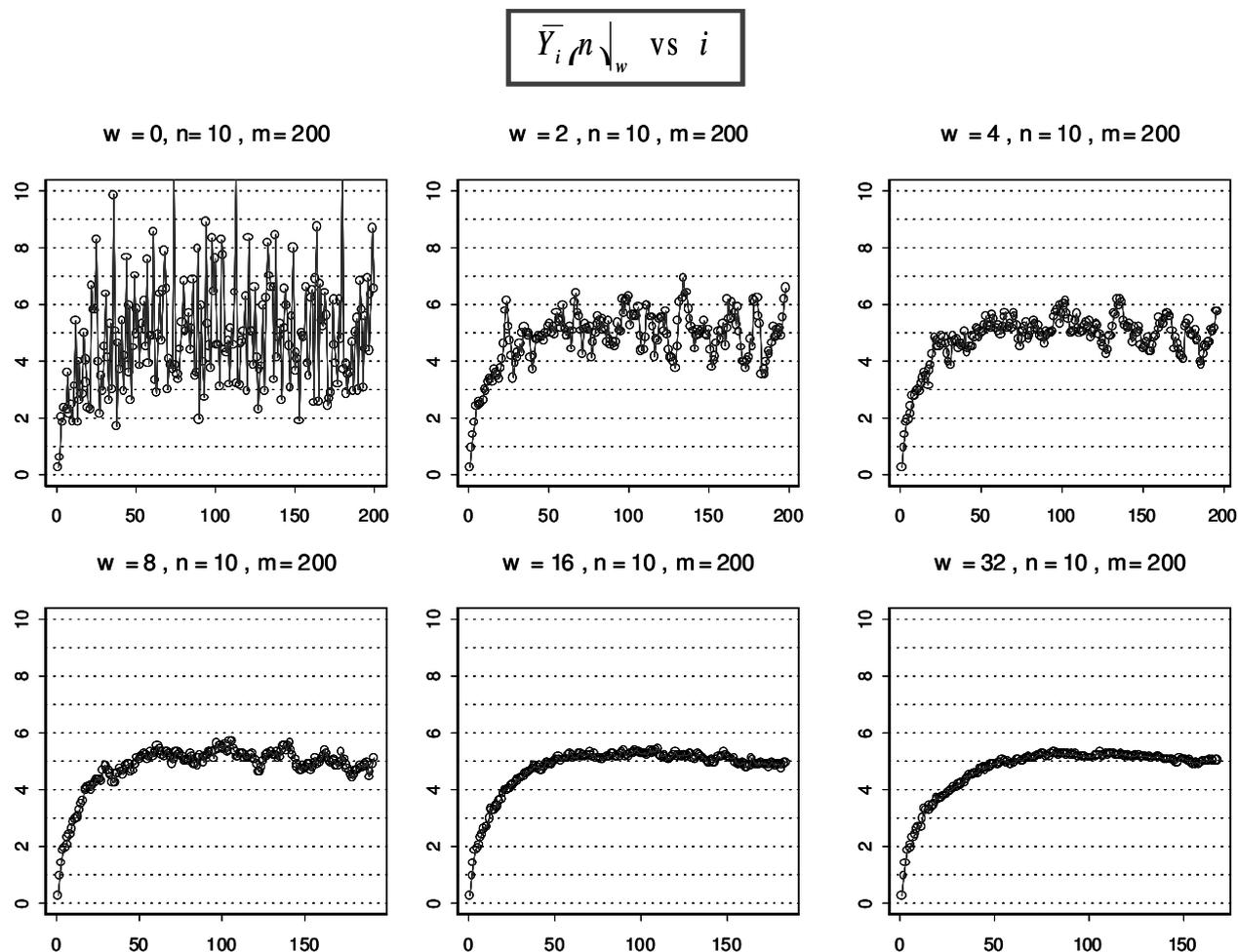


Figura 11.6: Aumentar la anchura de la ventana “suaviza” la curva.

Estimación del intervalo de confianza para la media de la distribución en el estacionario

Existen varias técnicas para estimar un intervalo de confianza para la media de la distribución en el estacionario. Puede establecerse una primera clasificación de estas técnicas atendiendo al número de réplicas de la simulación que requieran:

- Existen métodos basados en la realización de n réplicas independientes, de longitud m , de la simulación. Se discutirá uno de ellos: el *método de réplicas/eliminación*. Este método es aplicado en combinación con la técnica gráfica de Welch.
- Existen otras técnicas que están basadas en la realización de una única réplica “larga” de la simulación. En esta sección se describe una de ellas⁵: el *método de la media por lotes*, ya que es un método conceptualmente muy sencillo.

Las técnicas anteriores pueden aplicarse de acuerdo los dos enfoques siguientes:

- *Tamaño de muestra fijo*. Se obtiene una muestra de un determinado tamaño y se construye el intervalo de confianza usando estos datos.

⁵En el texto (Law & Kelton 2000) se discuten otras técnicas, como son el método auto-regresivo, el análisis espectral, el método regenerativo y el método de las series de tiempo estandarizadas.

- *Procedimiento secuencial.* Se va aumentando de forma secuencial el tamaño de la muestra, hasta que se obtiene un intervalo de confianza de longitud aceptable. Por ejemplo, en las técnicas basadas en una única réplica de la simulación, se va aumentando la longitud m de la simulación hasta que se obtiene un intervalo de confianza de longitud admisible.

Método de réplicas/eliminación

Se supone que, empleando la técnica gráfica de Welch (o cualquier otra igualmente válida), se ha obtenido una estimación de la longitud k del transitorio del proceso estocástico, $\{\bar{Y}_i(n) : i \geq 1\}$. Para ello, se han realizado n réplicas independientes de longitud m de la simulación (donde m es mucho mayor que k).

Eliminando las k primeras observaciones obtenidas de cada réplica de la simulación, y calculando la media de las restantes observaciones de cada réplica, pueden obtenerse las siguientes estimaciones (una por réplica) de la media de la distribución en el estacionario:

$$\bar{Y}_j(m, k) = \frac{1}{m - k} \cdot \sum_{i:k+1}^m y_{j,i} \quad \text{para } j : 1, 2, \dots, n \quad (11.18)$$

es decir, para cada réplica se calcula la media de las observaciones correspondientes al estado estacionario:

$$\begin{array}{ccccccc} y_{1,k+1} & y_{1,k+2} & \dots & y_{1,m} & \rightarrow & \bar{Y}_1(m, k) \\ \dots & & & & & \\ y_{n,k+1} & y_{n,k+2} & \dots & y_{n,m} & \rightarrow & \bar{Y}_n(m, k) \end{array} \quad (11.19)$$

Asumiendo que los estadísticos $\bar{Y}_1(m, k), \dots, \bar{Y}_n(m, k)$ son independientes y están idénticamente distribuidos, entonces:

$$\bar{Y}(n) = \frac{\bar{Y}_1(m, k) + \dots + \bar{Y}_n(m, k)}{n} \quad (11.20)$$

es un estimador aproximadamente no sesgado de μ_Y . Asimismo,

$$\bar{Y}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S^2(n)}{n}} \quad (11.21)$$

es un intervalo del $100 \cdot (1 - \alpha) \%$ de confianza para μ_Y . $S^2(n)$ es el estimador de la varianza de la muestra $\bar{Y}_1(m, k), \dots, \bar{Y}_n(m, k)$.

Método de las medias por lotes

A diferencia del método de réplicas/eliminación, en el que deben realizarse n réplicas de la simulación, el método de las medias por lotes se basa en la realización de una única simulación “larga”. La ventaja de este método es que sólo se simula el transitorio una vez, con lo cual el coste computacional para obtener observaciones “útiles” (observaciones en el estacionario) es menor que si se realizan n réplicas de la simulación (en cuyo caso se simula n veces el transitorio, en lugar de sólo una).

El método de las medias por lotes consta de los siguientes pasos:

1. Considérese el proceso estocástico Y_1, Y_2, \dots, Y_m , de covarianza estacionaria. Éste puede ser, o bien el proceso completo obtenido de la simulación, o bien el proceso con la parte transitoria eliminada. Se dividen las m observaciones en n grupos (denominados “lotes”) de igual longitud, $\frac{m}{n}$, y se calcula la media de las observaciones de cada lote:

$$\bar{Y}_i \left(\frac{m}{n} \right) = \frac{n}{m} \cdot \sum_{j:1}^{m/n} y_{(i-1) \cdot \frac{m}{n} + j} \tag{11.22}$$

Se obtiene:

$$\begin{array}{ccccccc} y_1 & y_2 & \dots & y_{\frac{m}{n}} & \rightarrow & \bar{Y}_1 \left(\frac{m}{n} \right) & \text{Lote 1} \\ y_{\frac{m}{n}+1} & y_{\frac{m}{n}+2} & \dots & y_{2 \cdot \frac{m}{n}} & \rightarrow & \bar{Y}_2 \left(\frac{m}{n} \right) & \text{Lote 2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ y_{(n-1) \cdot \frac{m}{n}+1} & y_{(n-1) \cdot \frac{m}{n}+2} & \dots & y_m & \rightarrow & \bar{Y}_n \left(\frac{m}{n} \right) & \text{Lote } n \end{array} \tag{11.23}$$

2. Si $\frac{m}{n}$ es suficientemente grande, entonces los estadísticos:

$$\bar{Y}_1 \left(\frac{m}{n} \right), \bar{Y}_2 \left(\frac{m}{n} \right), \dots, \bar{Y}_n \left(\frac{m}{n} \right) \tag{11.24}$$

son aproximadamente independientes, y están distribuidos aproximadamente de forma normal⁶. Así pues, si se representan $\bar{Y}(n)$ y $S^2(n)$ los estimadores de la media y la varianza de la muestra de n observaciones representada en la Expresión (11.24), entonces puede construirse un intervalo del $100 \cdot (1 - \alpha) \%$ de confianza para μ_Y de la forma siguiente:

$$\bar{Y}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S^2(n)}{n}} \tag{11.25}$$

11.4. Intervalos de confianza para varias magnitudes

En las anteriores secciones de este tema se ha discutido cómo construir un intervalo de confianza para una única magnitud. Sin embargo, en la mayoría de las aplicaciones prácticas es necesario conocer simultáneamente el comportamiento de varias magnitudes del sistema.

Supóngase que en un estudio de simulación (bien con condición de terminación, o bien en el estacionario) existen s magnitudes de interés:

$$\mu_1, \dots, \mu_s \tag{11.26}$$

y que se desea construir, para cada una de ellas, un intervalo de confianza. Sea I_i el intervalo del $100 \cdot (1 - \alpha_i) \%$ de confianza para μ_i , con $i : 1, \dots, s$. Pues bien, entonces sólo puede afirmarse que la probabilidad de que todos los intervalos de confianza contengan simultáneamente sus respectivos valores verdaderos satisface:

$$\text{Prob} \{ \mu_i \in I_i \text{ para todo } i : 1, \dots, s \} \geq 1 - \sum_{i:1}^s \alpha_i \quad (\text{Desigualdad de Bonferroti}) \tag{11.27}$$

sean o no I_i independientes. Esto tiene implicaciones serias para un estudio de simulación. En el siguiente ejemplo se ilustra este hecho.

Ejemplo 11.4.1. *Se construyen intervalos del 90% de confianza para 10 medidas diferentes del comportamiento del sistema. Entonces, únicamente se puede afirmar que:*

La probabilidad de que simultáneamente cada uno de los 10 intervalos contenga el valor verdadero de su magnitud, es mayor o igual que $1 - 10 \cdot 0.10 = 0$.

En consecuencia, no cabe tener mucha confianza en los resultados de este estudio. □

⁶Puede encontrarse la justificación a estas afirmaciones en el texto (Law & Kelton 2000).

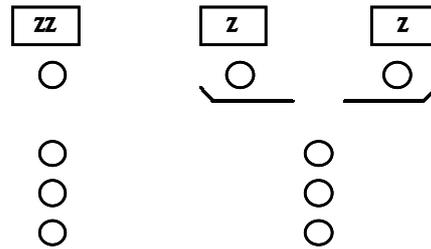


Figura 11.7: Comparación entre dos sistemas: ¿una máquina ZZ o dos Z?

Cuando el número de magnitudes de interés s es pequeño, una posible solución para que el nivel de confianza asociado conjuntamente a los s intervalos sea $100 \cdot (1 - \alpha) \%$, consiste en escoger los valores de α_i de modo que se satisfaga:

$$\sum_{i:1}^s \alpha_i = \alpha \quad (11.28)$$

Por ejemplo, se pueden construir 10 intervalos del 99% de confianza y tener una confianza global del 90%.

Obsérvese que los valores de α_i no tienen forzosamente que ser iguales. De hecho normalmente se emplean α_i menores para aquellas magnitudes más importantes.

11.5. Comparación entre dos sistemas

Hasta ahora las explicaciones se han centrado en cómo analizar los resultados obtenidos de la simulación de un único modelo. Sin embargo, en ocasiones el objetivo del estudio de simulación no es estimar el comportamiento de un único sistema, sino comparar el comportamiento de dos o más sistemas alternativos. El objetivo de este tipo de estudios puede ser comparar diferentes diseños, políticas de gestión, etc. A continuación se muestra un ejemplo de ello.

Ejemplo 11.5.1. *La empresa gestora de una estación de autobuses planea automatizar el proceso de venta de billetes a los viajeros. Se debe decidir si comprar dos máquinas expendedoras de billetes modelo Z o una máquina modelo ZZ. Las máquinas modelo ZZ tienen un precio de compra y mantenimiento doble a las Z, sin embargo las máquinas ZZ son el doble de rápidas que las Z. En caso de optarse por el modelo Z, se instalarían las dos máquinas una junto a otra y se formaría una única cola frente a ambas (ver Figura 11.7).*

Mediante un estudio de simulación, se trata de investigar cuál de las dos configuraciones (una máquina ZZ o dos Z) proporciona mejor servicio al cliente. Existen dos factores que pueden afectar a la satisfacción del cliente: el tiempo de espera en la cola y el tiempo de servicio.

Se realiza la siguiente hipótesis: el tiempo de espera en cola es, con mucha diferencia, el factor más importante. La espera en la cola es para los viajeros la parte más irritante del proceso de compra del billete. El tiempo de servicio afecta en mucha menor medida a la satisfacción del cliente, ya que durante el mismo éste percibe que está siendo atendido. En consecuencia, se decide ignorar en el estudio el tiempo de servicio al cliente como factor que contribuye a su satisfacción, considerándose únicamente el tiempo de espera en la cola.

Acerca de la estructura matemática del modelo, se realizan las aproximaciones siguientes:

- Los clientes llegan de uno en uno, de acuerdo a un proceso de Poisson con frecuencia 1 cliente/minuto.

- El tiempo de servicio de la máquina ZZ está distribuido exponencialmente, con media $\beta = 0.9$ minutos.
- El tiempo de servicio de cada máquina Z está distribuido exponencialmente, con media $\beta = 1.8$ minutos.

Se trata, por tanto, de comparar una cola M/M/1 con una cola M/M/2, cada una con un factor de utilización $\rho = 0.9$.

La medida de interés, que cuantifica el comportamiento del sistema, es el tiempo de espera medio en cola de los primeros 100 clientes que abandonan la cola. Se supone que inicialmente la cola está vacía y las máquinas libres. Dos formas inadecuadas de realizar el estudio de simulación son las siguientes:

- Una forma rudimentaria de plantear el estudio sería realizar un único experimento con cada sistema, consistente en una única simulación de longitud $m = 100$ clientes, y tomar la decisión en base al promedio de los tiempos de espera obtenidos en cada caso. Sin embargo, este procedimiento deja una pregunta sin respuesta: ¿cuál es la probabilidad de que este análisis proporcione la recomendación correcta?
- Un método algo menos rudimentario consiste en realizar 100 experimentos completos independientes, y anotar cuántas veces la simulación aconseja la instalación ZZ y cuántas la Z. Así, por ejemplo, si se obtiene que en 56 de los 100 experimentos el tiempo medio de espera del sistema ZZ era superior al del sistema Z, puede estimarse, con un 44% de probabilidad de error, que el sistema Z es preferible. Sin embargo, se ha fijado el número de experimentos de forma arbitraria, con lo cual cabe preguntarse si diferiría mucho el resultado obtenido si, en vez de haber realizado 100 experimentos, se hubieran realizado, por ejemplo, 1000.

Una vez más, la conclusión de todo ello es que es preciso emplear métodos estadísticos para realizar la comparación de manera rigurosa. \square

La comparación entre los dos sistemas se hace basándose en el valor esperado de determinada magnitud aleatoria en cada uno de ellos. Sean X_1 y X_2 las magnitudes de interés en los sistemas 1 y 2 respectivamente. Estas magnitudes son representativas del comportamiento del sistema a lo largo de toda la simulación: de cada réplica de la simulación se obtiene una única observación de la magnitud de interés.

Por ejemplo, X_1 y X_2 pueden ser los tiempos medios de espera por entidad en determinada cola del sistema 1 y 2 respectivamente. Para obtener una observación de X_i debe realizarse una réplica de la simulación del sistema i , y calcular el promedio de los tiempos obtenidos de espera de las entidades en la cola.

El objetivo del estudio es comparar el valor esperado de estas magnitudes de interés, es decir, comparar⁷ μ_1 con μ_2 . Por tanto, se trata de construir un intervalo de confianza para:

$$\mu_1 - \mu_2 \quad (11.29)$$

El intervalo de confianza no sólo responde a la pregunta de si la diferencia entre los valores esperados es significativa: es significativa en aquel caso en que el intervalo no contenga al cero. Además permite cuantificar esta diferencia. A continuación se describe un método para construir el intervalo de confianza.

1. Realizar n réplicas de la simulación de cada sistema, de modo que se obtengan n observaciones de la respuesta de interés:

$$\begin{array}{llll} x_{1,1} & \dots & x_{1,n} & n \text{ observaciones de } X_1 \\ x_{2,1} & \dots & x_{2,n} & n \text{ observaciones de } X_2 \end{array} \quad (11.30)$$

⁷Con el fin de simplificar la notación, en esta sección el valor esperado de X_i se representa μ_i , en lugar de μ_{X_i} , que resulta algo más engorroso.

2. Emparejar las observaciones anteriores y calcular la resta:

$$\begin{aligned} z_1 &= x_{1,1} - x_{2,1} \\ &\dots \\ z_n &= x_{1,n} - x_{2,n} \end{aligned} \tag{11.31}$$

No es preciso que $x_{1,i}$ sea independiente de $x_{2,i}$ (para $i : 1, \dots, n$).

3. Las n muestras de la variable aleatoria Z , definida como $Z = X_1 - X_2$, que han sido calculadas en el Paso 2, se emplean para construir un intervalo del $100 \cdot (1 - \alpha) \%$ de confianza para $\mu_Z = \mu_{X_1} - \mu_{X_2}$.

$$\bar{Z}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_Z^2(n)}{n}} \tag{11.32}$$

donde $\bar{Z}(n)$ y $S_Z^2(n)$ son los estimadores de la media y la varianza de las muestras z_1, \dots, z_n . Se calculan de la forma siguiente:

$$\bar{Z}(n) = \frac{\sum_{i:1}^n z_i}{n} \tag{11.33}$$

$$S_Z^2(n) = \frac{\sum_{i:1}^n (z_i - \bar{Z}(n))^2}{n - 1} \tag{11.34}$$

Si la variable aleatoria Z está distribuida de forma normal, $Z \sim N(\mu_Z, \sigma_Z^2)$, entonces el intervalo es exacto. En caso contrario, el intervalo es tanto más aproximado⁸ cuanto mayor sea n .

Ejemplo 11.5.2. La discusión anterior puede particularizarse al estudio del sistema descrito en el ejemplo 11.5.1. La magnitud de interés en cada sistema es el tiempo medio de espera en la cola, calculado de los tiempos de espera de las 100 primeras entidades. Se denomina X_i a la variable aleatoria siguiente: "tiempo medio de espera en cola en el sistema i ".

Deben realizarse n réplicas independientes de la simulación de cada sistema, con el fin de construir un intervalo de confianza para $\mu_1 - \mu_2$.

| | | |
|--|--|--|
| <p>Sistema 1</p> $\begin{array}{cccccc} d_{1,1,1} & d_{1,1,2} & \dots & d_{1,1,100} & \rightarrow & x_{1,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ d_{1,n,1} & d_{1,n,2} & \dots & d_{1,n,100} & \rightarrow & x_{1,n} \end{array}$ | | <p>Sistema 2</p> $\begin{array}{cccccc} d_{2,1,1} & d_{2,1,2} & \dots & d_{2,1,100} & \rightarrow & x_{2,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ d_{2,n,1} & d_{2,n,2} & \dots & d_{2,n,100} & \rightarrow & x_{2,n} \end{array}$ |
|--|--|--|

donde $d_{i,j,k}$ representa el tiempo que ha debido esperar el cliente k -ésimo, en la cola del sistema i , en la réplica número j de la simulación.

Emparejando las observaciones $x_{1,j}$ con las observaciones $x_{2,j}$, con $j : 1, \dots, n$, como se muestra en la Ecuación (11.31), se obtienen n muestras de la variable aleatoria $Z = X_1 - X_2$. Con ellas puede construirse un intervalo de confianza para el valor esperado de Z (ver la Ecuación (11.32)). □

⁸El teorema del límite central establece que:

Cuando el tamaño de la muestra, n , es grande, la media de la muestra $\bar{Z}(n)$ está aproximadamente distribuida como una variable aleatoria normal con media μ_Z y varianza $\frac{\sigma_Z^2}{n}$.

Esto se cumple sea cual sea la distribución de probabilidad de Z , lo cual justifica emplear el intervalo de confianza aproximado $\bar{Z}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_Z^2(n)}{n}}$ cuando n es grande.

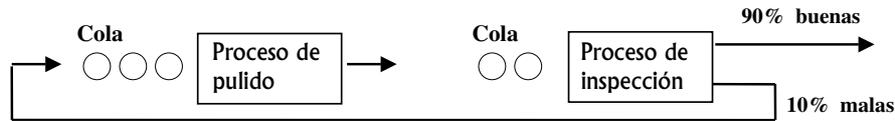


Figura 11.8: Esquema del funcionamiento de la línea de producción.

Ejemplo 11.5.3. Se desea comparar dos políticas de gestión del inventario (s, S) diferentes: $(20, 40)$ y $(20, 80)$. La magnitud a estudiar es el coste mensual, calculado promediando el coste de 120 meses de operación. La condición inicial es que hay inventariadas 60 unidades de producto.

Se realizan 10 réplicas independientes de la simulación de cada política de gestión. El coste mensual obtenido en cada réplica es el siguiente:

| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $x_{1,j}$ | 129.35 | 127.11 | 124.03 | 122.13 | 120.43 | 118.38 | 130.17 | 129.77 | 125.51 | 133.75 |
| $x_{2,j}$ | 127.39 | 122.52 | 123.08 | 119.19 | 124.46 | 118.04 | 123.67 | 121.59 | 127.38 | 119.13 |

Emparejando las observaciones, se obtiene el siguiente intervalo del 90% de confianza para la diferencia entre los valores esperados de los costes mensuales:

$$[0.30, 6.53] \quad (11.35)$$

Así pues, puede afirmarse con un 90% de confianza que el coste mensual de la política 1 es mayor que el de la política 2, con lo cual es preferible la política 2. \square

Ejemplo 11.5.4. Una pequeña línea de fabricación consiste en un proceso de pulido de piezas metálicas, y una estación de inspección en serie (ver Figura 11.8). Las piezas llegan a la fábrica, son pulidas y a continuación son inspeccionadas. Las piezas bien pulidas (“buenas”) son enviadas y las mal pulidas (“malas”) son vueltas a pulir. De cuando en cuando la pulidora se avería y durante el tiempo en que está siendo reparada el proceso de pulido no puede realizarse.

Los aspectos matemáticos del modelo son los siguientes:

- Los intervalos de tiempo entre la llegada de piezas están distribuidos exponencialmente, con media 1 minuto.
- El tiempo del proceso de pulido está distribuido uniformemente, en el intervalo $[0.65, 0.70]$ minutos.
- El tiempo del proceso de inspección está uniformemente distribuido en el intervalo $[0.75, 0.80]$ minutos.
- El 90% de las piezas inspeccionadas están bien pulidas. Debe repetirse el pulido del 10% restante.
- El intervalo de tiempo entre averías de la pulidora está distribuido exponencialmente, con media 360 minutos.
- El tiempo necesario para reparar la pulidora está distribuido uniformemente en el intervalo $[8, 12]$ minutos.

Se está estudiando sustituir el equipo de inspección actual por uno nuevo. El nuevo equipo reducirá el tiempo de inspección en un 10%, de modo que éste quedaría distribuido uniformemente en el intervalo $[0.675, 0.720]$.

El objetivo del estudio de simulación es determinar si, una vez que el funcionamiento de la fábrica ha alcanzado el estacionario, este cambio reduce significativamente el tiempo medio de fabricación de las piezas. El tiempo de fabricación de una pieza es el tiempo que transcurre entre que se recibe en la fábrica hasta que se envía.

Se denomina $T_{i,p}$ a la variable aleatoria “tiempo de fabricación de la pieza p en el proceso i ”, donde $i = 1$ es el proceso actual e $i = 2$ es el proceso nuevo.

El estudio consta de los pasos siguientes:

1. Debe estimarse la longitud del periodo transitorio de cada uno de los dos sistemas, k_1 y k_2 , empleando la técnica gráfica de Welch. Para ello, es preciso realizar n réplicas independientes de longitud m_1 de la simulación del sistema 1, y otras n réplicas de longitud m_2 de la simulación del sistema 2.

| | |
|--|--|
| Sistema 1 | Sistema 2 |
| $t_{1,1,1} \quad t_{1,1,2} \quad \dots \quad t_{1,1,m_1}$ $\vdots \quad \quad \quad \ddots \quad \quad \quad \vdots$ $t_{1,n,1} \quad t_{1,n,2} \quad \dots \quad t_{1,n,m_1}$ | $t_{2,1,1} \quad t_{2,1,2} \quad \dots \quad t_{2,1,m_2}$ $\vdots \quad \quad \quad \ddots \quad \quad \quad \vdots$ $t_{2,n,1} \quad t_{2,n,2} \quad \dots \quad t_{2,n,m_2}$ |

2. Se calcula el tiempo medio de proceso en el estacionario de cada réplica. Para ello, no se consideran las k_i primeras observaciones de cada réplica.

| | | | |
|------------|--|---------------|--|
| Sistema 1: | $t_{1,1,1} \quad t_{1,1,2} \quad \dots \quad t_{1,1,m_1}$ $\vdots \quad \quad \quad \ddots \quad \quad \quad \vdots$ $t_{1,n,1} \quad t_{1,n,2} \quad \dots \quad t_{1,n,m_1}$ | \rightarrow | $z_{1,1} = \frac{\sum_{p:k_1+1}^{m_1} t_{1,1,p}}{m_1 - k_1}$ \vdots $z_{1,n} = \frac{\sum_{p:k_1+1}^{m_1} t_{1,n,p}}{m_1 - k_1}$ |
| Sistema 2: | $t_{2,1,1} \quad t_{2,1,2} \quad \dots \quad t_{2,1,m_2}$ $\vdots \quad \quad \quad \ddots \quad \quad \quad \vdots$ $t_{2,n,1} \quad t_{2,n,2} \quad \dots \quad t_{2,n,m_2}$ | \rightarrow | $z_{2,1} = \frac{\sum_{p:k_2+1}^{m_2} T_{1,1,p}}{m_2 - k_2}$ \vdots $z_{2,n} = \frac{\sum_{p:k_2+1}^{m_2} T_{1,n,p}}{m_2 - k_2}$ |

3. Se realizan los emparejamientos de las muestras anteriores, $z_{1,j}$ y $z_{2,j}$, para obtener las n observaciones siguientes:

$$z_j = z_{1,j} - z_{2,j} \quad \text{para } j : 1, \dots, n \tag{11.36}$$

En el texto (Law & Kelton 2000) se muestra el estudio de simulación completo. Se aplica la técnica gráfica del Welch a $n = 20$ réplicas de la simulación y se escogen los valores siguientes:

$$\begin{aligned} k_1 &= 2286 & m_1 &= 9407 \\ k_2 &= 2093 & m_2 &= 9434 \end{aligned}$$

Se obtiene el intervalo del 90% de confianza siguiente:

$$2.36 \pm 0.31 \tag{11.37}$$

Así pues, la diferencia entre la media de los tiempos de proceso en el estacionario de los sistemas 1 y 2 es estadísticamente significativa. Representa una reducción aproximada del 38% (3.8 vs 6.6). Obsérvese que esta reducción del 38% en el tiempo medio de proceso en el estacionario se ha conseguido mediante una reducción únicamente del 10% en el tiempo medio de proceso de la inspección.

□

11.6. Comparación entre más de dos sistemas

En ocasiones, el objetivo de un estudio de simulación es realizar una comparación entre k sistemas (con $k > 2$). Puede realizarse comparando los k sistemas entre sí, o bien comparando cada uno de ellos con un determinado sistema estándar.

La comparación se realiza en base al valor esperado de determinada magnitud de interés de cada sistema. Sea X_i la magnitud de interés del sistema i y sea $\mu_i = E\{X_i\}$. Tal como se indicó en la sección anterior, estas magnitudes de interés debe ser representativas del comportamiento del sistema durante la réplica de la simulación.

El procedimiento para establecer las comparaciones suele ser construir un determinado número, s , de intervalos de confianza:

$$I_1, \dots, I_s \quad (11.38)$$

donde I_i es un intervalo del $100 \cdot (1 - \alpha_i)$ % de confianza para $\mu_i = E\{X_i\}$. El nivel de cada uno de estos intervalos I_i debe ajustarse de acuerdo a la *confianza global* deseada, $100 \cdot (1 - \alpha)$ %. Este nivel de confianza global debe interpretarse de la manera siguiente:

La probabilidad del evento $\{\mu_1 \in I_1, \dots, \mu_s \in I_s\}$ es al menos $1 - \alpha$.

La relación matemática entre el nivel de confianza global y el nivel de confianza de cada uno de los s intervalos es consecuencia de la desigualdad de Bonferroni (ver la Ecuación (11.27)):

$$\alpha = \sum_{i:1}^s \alpha_i \quad (11.39)$$

La consecuencia práctica es que si el número s de intervalos es elevado y se desea obtener un valor de α razonablemente pequeño, entonces los valores de $\alpha_1, \dots, \alpha_s$ deberán ser muy pequeños y, en consecuencia, los intervalos I_1, \dots, I_s serán considerablemente anchos. Para reducir la anchura de los mismos, deberá aumentarse el tamaño de la muestra.

Comparación con un estándar

En este caso, el primero de los k sistemas es el estándar, respecto al cual se desean comparar los $k - 1$ sistemas restantes. Para realizar la comparación, ha de construirse el intervalo de confianza para cada una de las $s = k - 1$ diferencias siguientes:

$$\mu_2 - \mu_1, \dots, \mu_k - \mu_1 \quad (11.40)$$

de modo que el nivel de confianza *global* de los s intervalos sea $1 - \alpha$. El intervalo de confianza para $\mu_i - \mu_1$, con $i : 2, \dots, k$, se representa I_{i-1} .

Una vez construidos los intervalos, se puede afirmar con un nivel de confianza de al menos $1 - \alpha$ que:

- el sistema i difiere del sistema estándar si el intervalo para $\mu_i - \mu_1$ no contiene el cero, y que
- el sistema i no es significativamente diferente del estándar si el intervalo para $\mu_i - \mu_1$ contiene el cero.

Dado que la desigualdad de Bonferroni es muy general (Ecuación (11.27)), no importa cómo se hayan construido los intervalos de confianza, es decir,

- El número de réplicas de la simulación, usado para construir cada intervalo de confianza, puede ser diferente.

- No es necesario que las réplicas para la construcción de un intervalo sean independientes de las réplicas para la construcción de otro.

Comparación de todos con todos

En algunos estudios, en los que se analizan k sistemas, es preciso comparar cada sistema con los $k - 1$ restantes⁹. Un procedimiento consiste en construir intervalos de confianza para las $s = \frac{k \cdot (k-1)}{2}$ diferencias siguientes:

$$\mu_i - \mu_j \quad \text{para} \quad \begin{cases} i : 1, \dots, k-1 \\ j : i+1, \dots, k \end{cases} \quad (11.41)$$

Ejemplo 11.6.1. Para comparar $k=5$ sistemas entre sí, deberían construirse los siguientes 10 intervalos:

| | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ |
|---------|-----------------|-----------------|-----------------|-----------------|
| $i = 1$ | $\mu_1 - \mu_2$ | $\mu_1 - \mu_3$ | $\mu_1 - \mu_4$ | $\mu_1 - \mu_5$ |
| $i = 2$ | | $\mu_2 - \mu_3$ | $\mu_2 - \mu_4$ | $\mu_2 - \mu_5$ |
| $i = 3$ | | | $\mu_3 - \mu_4$ | $\mu_3 - \mu_5$ |
| $i = 4$ | | | | $\mu_4 - \mu_5$ |

□

La relación que existe entre el nivel de confianza de los intervalos y el nivel de confianza global es: $\alpha = \sum_{i:1}^s \alpha_i$.

11.7. Análisis usando Arena

En esta sección se discute cómo interpretar los resultados estadísticos que ofrece Arena al realizar la simulación de un modelo. Tanto en el caso *con terminación*, en el que se realizan varias réplicas independientes de la simulación, como en el caso *estacionario*, en el que se hace una única réplica “larga”. Seguidamente, se describirá cómo emplear la herramienta *Arena Output Analyzer* para comparar dos versiones alternativas de un modelo.

Simulación con terminación de un modelo

En el caso con terminación, la forma de obtener n observaciones de las respuestas es realizar n réplicas independientes de la simulación. El número de réplicas se indica al definir el experimento (pulsando *Run / Setup ...*), en la casilla “Number of Replications”. Para obtener réplicas independientes e idénticamente distribuidas de la simulación, es preciso marcar las casillas “Statistics” y “System”, de modo que el estado del sistema y los estadísticos se inicialice entre réplicas, y además Arena emplee secuencias de números aleatorios separadas para cada réplica.

⁹En (Law & Kelton 2000) se discuten algunas técnicas estadísticas adecuadas para otros problemas relacionados con este, como son:

- Selección del mejor de k sistemas.
- La selección de un subconjunto de m sistemas que contenga al mejor de los k sistemas. La utilidad de este problema es servir de paso previo a la selección del sistema óptimo cuando el número k de sistemas es grande.
- La selección de los m mejores sistemas de un conjunto de k .

| Statistic - Advanced Process | | | | | |
|------------------------------|--------------------|--------|----------------------|--------------------|-------------|
| | Name | Type | Expression | Report Label | Output File |
| 1 | llamadas atendidas | Output | EntitiesOut(llamada) | llamadas atendidas | llamada.dat |

Double-click here to add a new row.

Figura 11.9: Definición de un estadístico en Arena y volcado a fichero.

Finalizadas las n réplicas¹⁰, Arena vuelca en el fichero `.out` el resultado obtenido en cada réplica. Además, a partir del valor de cada magnitud obtenido en cada réplica, calcula el valor medio de las n réplicas y el intervalo del 95% de confianza para la media. Las magnitudes para las que Arena realiza este cálculo son clasificadas en el fichero `.out` como “OUTPUTS”. A continuación de muestra un ejemplo.

Ejemplo 11.7.1. Se realizan 10 réplicas de la simulación de un modelo en el que se define la entidad “llamada”. En el fichero “.out” se obtiene el número de entidades que han abandonado el sistema en cada réplica, “llamada.NumberOut”, y además el intervalo de confianza para la media calculado a partir de estas 10 observaciones. En las siguientes tablas se muestran los resultados:

| Réplica | llamada.NumberOut |
|---------|-------------------|
| 1 | 736.00 |
| 2 | 707.00 |
| 3 | 713.00 |
| 4 | 756.00 |
| 5 | 697.00 |
| 6 | 771.00 |
| 7 | 703.00 |
| 8 | 678.00 |
| 9 | 729.00 |
| 10 | 711.00 |

| Identifier | Average | Half-width | Minimum | Maximum | # Replications |
|-------------------|---------|------------|---------|---------|----------------|
| llamada.NumberOut | 720.10 | 20.112 | 678.00 | 771.00 | 10 |

□

En el módulo de datos “Statistic” (panel “Advanced Process”) pueden definirse estadísticos que sean interés para el estudio de simulación. Si el estadístico se define del tipo (“Type”) “Output”, entonces en cada réplica de la simulación se vuelca el valor del estadístico al fichero “.out” (dentro de la sección “OUTPUTS”), y finalizadas las n réplicas se calcula el intervalo del 95% de confianza para la media, usando para ello las n observaciones obtenidas (una por réplica). Si además se desea que Arena escriba estos n valores en un fichero binario (por ejemplo, con el fin de estudiarlos empleando *Arena Output Analyzer*), entonces debe especificarse el nombre del fichero en la casilla de la columna “Output File”, indicando explícitamente la extensión, que por convenio debe ser “.dat”.

Ejemplo 11.7.2. Continuando con el ejemplo 11.7.1, en la Figura 11.9 se muestra la definición del estadístico “llamadas atendidas”. Su valor es “EntitiesOut(llamada)”, es decir, es igual a “llamada.NumberOut”. El valor que adquiere este nuevo estadístico en cada réplica de la simulación se vuelca al fichero “.out”, y finalizadas las réplicas se calcula el intervalo del 95% de confianza para su media. Además, Arena genera el archivo “llamada.dat”, en el que se vuelcan las n observaciones del estadístico. □

¹⁰Con el fin de acelerar la ejecución de las simulaciones, suele deshabilitarse la animación cuando se realizan varias réplicas de la simulación. Para ello, hay que entrar en *itsshape Run / Run Control* y marcar *Batch Run (No Animation)*.

Arena calcula el intervalo de confianza aplicando la expresión (11.7). Asume, por tanto, que las n observaciones son independientes y que están idénticamente distribuidas de forma normal. Esta suposición está bastante justificada (por el teorema del límite central) cuando el valor que se obtiene de cada réplica es la suma o el promedio de otras magnitudes, por ejemplo, el tiempo medio en una cola, el tiempo medio de la acción "Delay" de un proceso, el tiempo de ciclo medio, el número de entidades que abandonan el sistema, el coste total acumulado, etc. Por el contrario, la hipótesis de normalidad no estaría a priori justificada si el valor que se obtiene de cada réplica de la simulación es un máximo o un mínimo: el tiempo máximo en una cola, en una determinada acción "Delay", el tamaño máximo de los grupos de entidades que han llegado al sistema, etc.

Comparación de dos modelos

La herramienta Arena Output Analyzer¹¹ permite comparar la respuesta de dos modelos. Se trata de un programa que se ejecuta independientemente de Arena, pero que opera sobre los ficheros creados por Arena a través del módulo de datos "Statistic" (los ficheros ".dat" anteriormente mencionados).

Los ficheros de entrada a Output Analyzer se seleccionan mediante *File / New*. No sólo se selecciona un único fichero ".dat", sino todos los ficheros ".dat" que contienen datos de interés para el estudio. En la ventana de diálogo que se abre, se van añadiendo los ficheros ".dat" uno a uno. Una vez se han seleccionado todos, se cierra la ventana pulsando el aspa en la parte superior derecha. Entonces, Output Analyzer pregunta si debe guardar los nombres de estos ficheros ".dat" en un fichero. Si se escoge que sí los guarde, la herramienta almacena la información en un fichero con extensión ".dgr", que podrá abrirse posteriormente pulsando *File / Open*. De este modo se evita tener que volver a seleccionar los nombres de los ficheros ".dat".

La comparación puede realizarse seleccionando *Analyze / Compare Means*. Se abre una ventana en la cual debe indicarse:

- Las parejas de ficheros ".dat" a comparar: A y B .
- El nivel de confianza del test: 0.90, 0.95 ó 0.99.
- Debe escogerse entre "Paired-t Test" (la opción por defecto) y "Two-Sample-t Test".

Output Analyzer calcula la resta de las medias en la dirección $A - B$, mostrando el intervalo de confianza para el valor medio de la diferencia. Si el intervalo no contiene el cero, entonces existe evidencia de que existen diferencias estadísticamente significativas entre los dos modelos. Asimismo, se muestran los resultados del test estadístico cuya hipótesis nula es que no existe diferencia entre las medias. Obsérvese que el intervalo de confianza proporciona una información más completa que el test, ya que no sólo indica si existe o no diferencia, sino de que además cuantifica esta diferencia.

Ejemplo 11.7.3. *Se pretenden comparar dos configuraciones de un sistema de atención telefónica. La magnitud que mide el rendimiento del sistema es el número total de llamadas atendidas al día: cuanto mayor sea este número, mejor. Se realizan 10 réplicas de la simulación de cada sistema, y se vuelcan los resultados obtenidos con la configuración actual del sistema al fichero "llamada.dat", y los resultados obtenidos de la configuración alternativa al fichero "llamada1.dat".*

Se emplea Output Analyzer para comparar estadísticamente el valor esperado del número de llamadas atendidas diariamente por cada configuración. Pulsando "File / New" se seleccionan los dos ficheros ".dat", y seguidamente se pulsa "Analyze / Compare Means". En la Figura 11.10 se muestra la forma en que son introducidos los datos, y en la Figura 11.11 el resultado de la comparación. Puede observarse que existen diferencias estadísticamente significativas entre las dos configuraciones: la configuración B (fichero llamada1.dat) permite atender diariamente aproximadamente 750 llamadas más. □

¹¹ Arena Output Analyzer se suministra en el mismo CD que Arena, pero no es instalado por la instalación típica por defecto. Hay que seleccionar la instalación "Custom", y entonces indicar que se desea instalar.

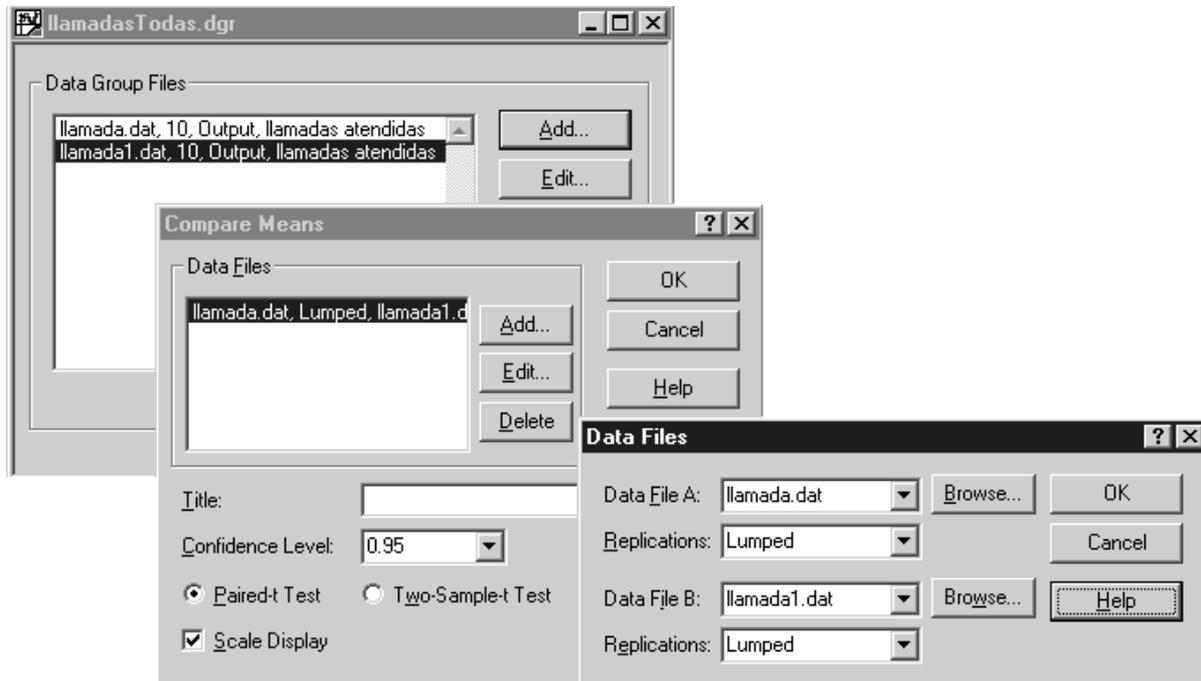


Figura 11.10: Empleo del *Output Analyzer* para comparar las medias.

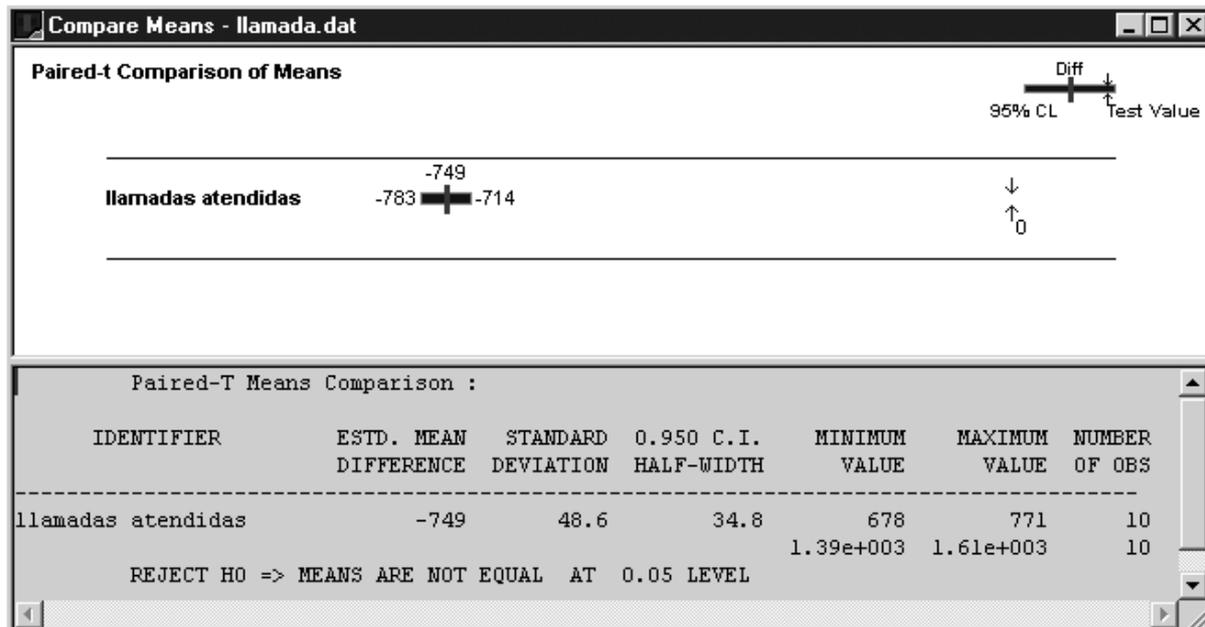


Figura 11.11: Comparación entre las medias.

Comparación entre más de dos modelos

Para comparar estadísticamente el comportamiento de varios modelos es preciso, en primer lugar realizar n réplicas de la simulación de cada modelo, y en segundo lugar evaluar los resultados estadísticamente. La herramienta *Process Analyzer* (abreviado PAN) es una aplicación (distribuida junto con Arena) que facilita la realización de estas dos tareas. PAN puede arrancarse desde Arena, pulsando “Tools / Process Analyzer”.

PAN permite establecer comparaciones estadísticas entre distintos escenarios. Un *escenario* (“Scenario”) está compuesto por:

- El fichero con extensión “.p”, en el que se encuentra la definición del modelo. Este tipo de fichero lo genera Arena al ejecutar la simulación y también al comprobar el modelo mediante “Run / Check Model”. Éste es el fichero sobre el que opera PAN.
- Un conjunto de valores para las *entradas de control* (“input controls”). Estos *controles* se escogen entre las variables del modelo y las capacidades de sus recursos.
- El conjunto de respuestas (“output responses”), sobre las cuales se desea establecer la comparación. PAN permite escoger las respuestas entre todos los estadísticos que Arena vuelca al fichero “.out” cuando se simula el modelo.

Una vez definidos los escenarios que se desean comparar, PAN ejecuta las réplicas de la simulación de cada escenario, almacenando los valores obtenidos para las respuestas en una tabla. Para usar PAN eficientemente, es preciso pensar, al definir el modelo en Arena, qué parámetros van a querer modificarse en el estudio, y definirlos como variables del modelo (o capacidades de los recursos), de modo que puedan ser seleccionados como entradas de control de los escenarios.

Una vez arrancado PAN, para empezar a trabajar hay que crear un nuevo *proyecto* (“File / New”), o abrir uno ya existente (“File / Open”). El *proyecto* contiene la definición de los escenarios con los que se desea experimentar. Si se escoge crear un nuevo proyecto, éste inicialmente se encuentra vacío; deben definirse sus escenarios. La definición de un *escenario* consta de los tres pasos siguientes:

1. *Propiedades del escenario*. Para definir un nuevo escenario hay que hacer doble “clic” donde se indica: “Double-click here to add a new scenario”. Al hacer doble “clic”, se abre una ventana con la cabecera “Scenario Properties”, en la que se debe indicar el nombre del escenario, y el fichero “.p” (que ya debe estar creado) que contiene el modelo. Para editar posteriormente estas propiedades, hay que pulsar con el botón derecho del ratón sobre la línea del escenario y seleccionar “Scenario Properties”.
2. *Entradas de control*. Una vez definido el escenario, pueden indicarse cuáles son sus entradas de control. Para ello hay que pulsar con el botón derecho del ratón sobre la línea del escenario y seleccionar “Insert Control”. Se abre una ventana en la que se muestran las posibles elecciones, clasificadas en las tres categorías siguientes:
 - Los recursos del modelo, cuya capacidad puede escogerse como entrada de control.
 - Las variables definidas por el usuario en el modelo.
 - Las variables del sistema: longitud y número de réplicas.

Cada una de las categorías se desdobra en forma de árbol. Haciendo doble “clic” sobre las diferentes opciones se van seleccionando las entradas de control. Las entradas de control escogidas se muestran en la fila correspondiente al escenario, bajo la cabecera “Control”. El valor que se muestra para la entrada de control es el definido en el modelo en Arena, pudiendo ahora ser modificado si se desea.

3. *Respuestas*. Una vez seleccionadas las entradas de control, puede procederse a seleccionar las respuestas. Para ello, haciendo “clic” con el botón derecho del ratón sobre la fila del escenario, hay que pulsar “Insert Response”. Se muestran aquellas magnitudes del modelo candidatas a ser respuestas. Para seleccionar una de estas magnitudes como respuesta, hay que hacer doble “clic” sobre ella. Entonces, es añadida en la fila del escenario, sin ningún valor por el momento, ya que todavía no se ha ejecutado el escenario.

Usando este procedimiento, pueden definirse los escenarios que se desee comparar. Para crear un escenario similar a otro ya definido, resulta más cómodo hacer una copia de éste y hacer sobre ella las modificaciones precisas. Para hacer una copia de un escenario, debe pulsarse con el botón derecho del ratón sobre el número del escenario (primera columna) y seleccionar "Duplicate Scenario".

Ejemplo 11.7.4. *Se ha construido un modelo de un sistema de atención telefónica, con el fin de estudiar el impacto que tiene el número de líneas telefónicas sobre el número de "llamadas perdidas". Los clientes llaman a un número único, que es atendido por 10 líneas telefónicas. Si un cliente llama en un instante en el que todas las líneas se encuentran ocupadas, entonces recibe la señal de comunicando y debe colgar: se pierde una llamada. Los detalles internos de funcionamiento del sistema no son relevantes para este ejemplo. Al comprobar el modelo del sistema ("Run / Check Model"), que se ha llamado "modelo.doe", Arena genera el fichero "modelo.p".*

Se desea estudiar cómo afecta la capacidad del recurso "línea telefónica" al número de llamadas perdidas (variable "Llamadas perdidas"). La capacidad del recurso es igual al número de líneas que dan servicio al número telefónico central. El estudio se realiza definiendo los cinco escenarios siguientes: la situación actual del sistema, con 10 líneas, y cuatro configuraciones alternativas: 11, 12, 13 y 14 líneas telefónicas. En la Figura 11.12 se muestra la definición de los 5 escenarios. Se han definido dos variables de control: el número de líneas telefónicas y el número de réplicas de la simulación.

Para ejecutar un subconjunto de los escenarios, o todos ellos, hay que seleccionar sus filas haciendo "clic" en la correspondiente casilla de la columna situada más a la izquierda, en la que se muestran los números de fila. A continuación, hay que pulsar Run / Go. Se abre una ventana mostrando los escenarios escogidos y solicitando confirmación. Pulsando OK se inicia la ejecución de las 50 réplicas de cada escenario. Los resultados obtenidos se muestran en la Figura 11.13. El valor mostrado, para la respuesta en cada escenario, corresponde con el valor medio calculado de las 50 observaciones, obtenidas de las 50 réplicas de la simulación.

Puede observarse en qué medida el número medio de llamadas perdidas decrece al incrementarse el número de líneas. Sin embargo, observar únicamente los valores medios da una información muy pobre. Es posible representar gráficamente el boxplot de las 50 observaciones de cada escenario. Para ello, debe seleccionarse (pulsando con el botón izquierdo) la columna con la respuesta y, a continuación, pulsando el botón derecho, seleccionar "Insert Chart". Para representar los boxplots, el tipo de gráfico debe ser "Box and Whisker".

Una vez dibujado el gráfico, pulsando sobre él con el botón derecho del ratón, y seleccionando "Chart Options", pueden verse los valores numéricos a partir de los cuales se han dibujado los Boxplots. Estos son:

| # líneas | Min | Max | Lo | Hi | 95 % CI |
|----------|-----|-----|-------|-------|---------|
| 10.0000 | 315 | 429 | 357.8 | 372.5 | 7.361 |
| 11.0000 | 294 | 407 | 325.3 | 338.9 | 6.829 |
| 12.0000 | 259 | 367 | 295.3 | 309.9 | 7.303 |
| 13.0000 | 212 | 335 | 265.2 | 278.6 | 6.744 |
| 14.0000 | 213 | 318 | 243.5 | 255.1 | 5.806 |

PAN dibuja los Boxplots de una forma diferente a la que se ha venido empleando en los temas anteriores. La barra vertical representa el rango de los valores, es decir, va desde el valor mínimo (Min en la tabla anterior) hasta el valor máximo (Max en la tabla). La caja representa el intervalo del 95% de confianza, que va desde el valor Lo de la tabla hasta el Hi. La mitad de la anchura de este intervalo, $\frac{Hi-Lo}{2}$, es el valor señalado en la tabla como 95 % CI (CI significa "Confidence Interval"). □

Simulación en el estacionario

La forma más habitual de realizar análisis en el estacionario usando Arena es inicializar el sistema de modo que las colas se encuentren vacías y los recursos libres. Dejar avanzar la

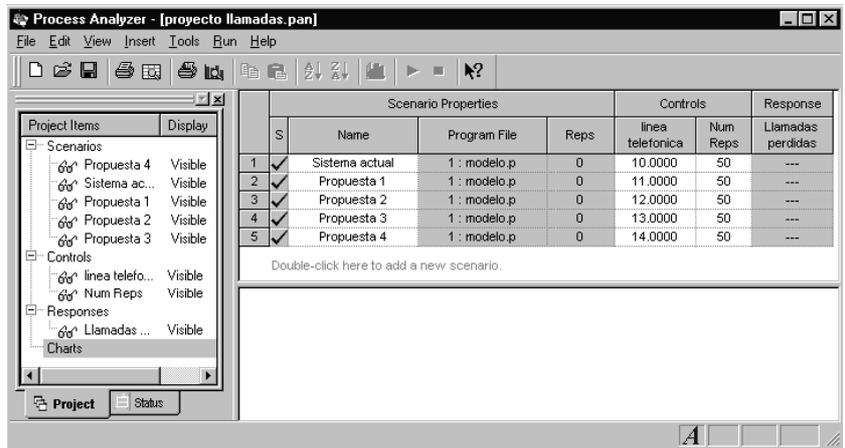


Figura 11.12: Definición de escenarios en PAN.

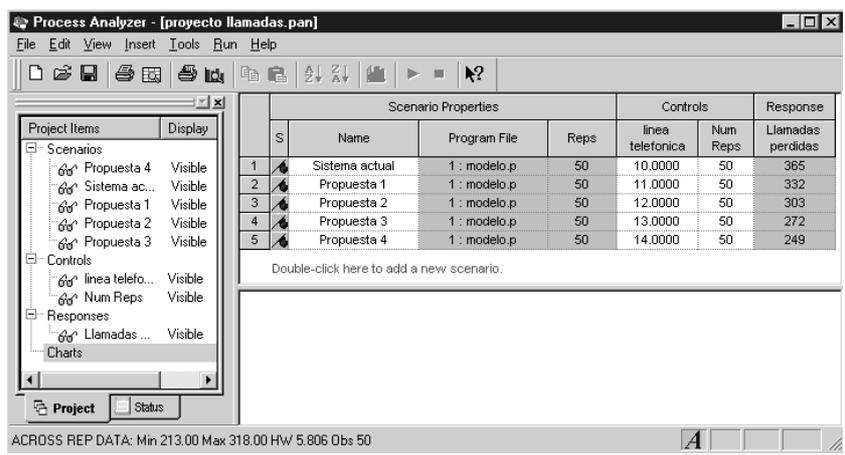


Figura 11.13: Resultado de las simulaciones.

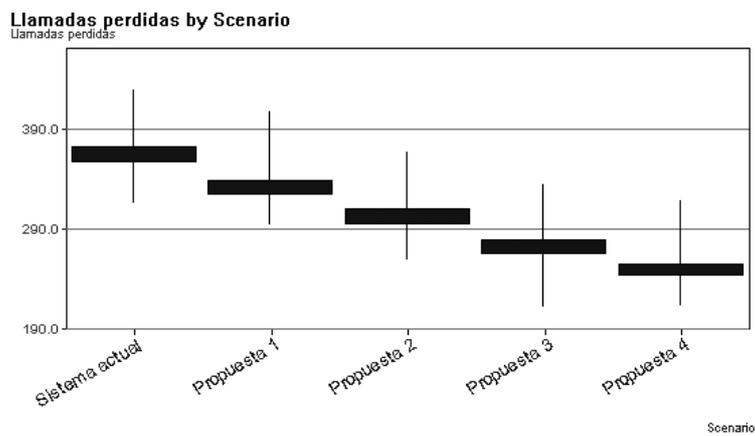


Figura 11.14: Boxplots de las 50 observaciones de cada escenario.

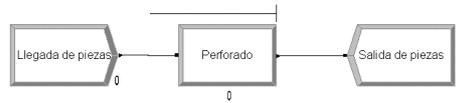


Figura 11.15: Diagrama de módulos del proceso de perforado.

| Statistic - Advanced Process | | | | | |
|------------------------------|------------|-----------------|-------------------------------------|--------------|----------------|
| | Name | Type | Expression | Report Label | Output File |
| 1 | piezasCola | Time-Persistent | DAVG(Perforado.Queue.NumberInQueue) | piezasCola | piezasCola.dat |

Double-click here to add a new row.

Figura 11.16: Definición del estadístico.

simulación durante cierto tiempo, hasta que el comportamiento del sistema alcance aproximadamente el estacionario. En ese instante, una vez transcurrido el transitorio (“warm up”), los contadores estadísticos son inicializados y comienzan a recogerse datos de la simulación. De este modo, el informe de la simulación únicamente refleja lo que sucede una vez pasado el periodo transitorio. La duración del periodo transitorio (“Warm-up Period”) se especifica al definir el experimento (“Run / Setup / Replication Parameters”). No hay que olvidar especificar las unidades de tiempo (“Time Units”) en que se expresa este periodo.

Para establecer la duración del periodo transitorio, puede resultar útil salvar en un fichero la evolución de la respuesta a lo largo de la simulación. Para ello debe definirse respuesta en el módulo de datos “Statistic”, del panel “Advanced Process”, especificando la expresión a partir de la cual debe calcularse, y el nombre del fichero en el que deben guardarse los datos. A continuación se muestra un ejemplo.

Ejemplo 11.7.5. *Considérese el proceso de perforado de piezas descrito en el Problema 6.1. El diagrama de módulos se muestra en la Figura 11.15. Se desea analizar el funcionamiento del sistema en el estado estacionario, más concretamente, el número medio de piezas en cola del proceso de perforado. Se define el estadístico (piezasCola, del tipo “Time-persistent”, según se muestra en la Figura 11.16, especificando el nombre del fichero al que Arena debe volcar los valores.*

A continuación se define el experimento (“Run / Setup”). El periodo transitorio se deja a cero (valor por defecto), se asigna un valor muy grande a la duración de la simulación (“Replication Length”) y se establece el número de réplicas. En este caso, se realizan 10 réplicas de 4000 minutos de duración cada una.

Finalizadas las 10 réplicas, Arena ha creado el fichero “piezasCola.dat” en el directorio donde se encuentra el modelo. Para visualizar su contenido es preciso arrancar Output Analyzer, y crear una nueva sesión (“File / New”). Se abre una ventana (“Data Group Files”), en la que deben añadirse (“Add”) los ficheros con los datos. En este caso, hay que añadir el fichero “piezasCola.dat”.

En la Figura 11.17 se muestra la evolución (“Graph / Plot”) del número medio de piezas en cola en cada una de las 10 réplicas de la simulación. Puede observarse que aproximadamente a partir del minuto 2000 alcanza el estacionario. Siendo conservadores, puede suponerse que el periodo transitorio del proceso dura aproximadamente 3000 minutos, en lugar de 2000 minutos. Para realizar el estudio en el estacionario, debe indicarse este valor del periodo transitorio (“Warm-up Period”) al definir el experimento.

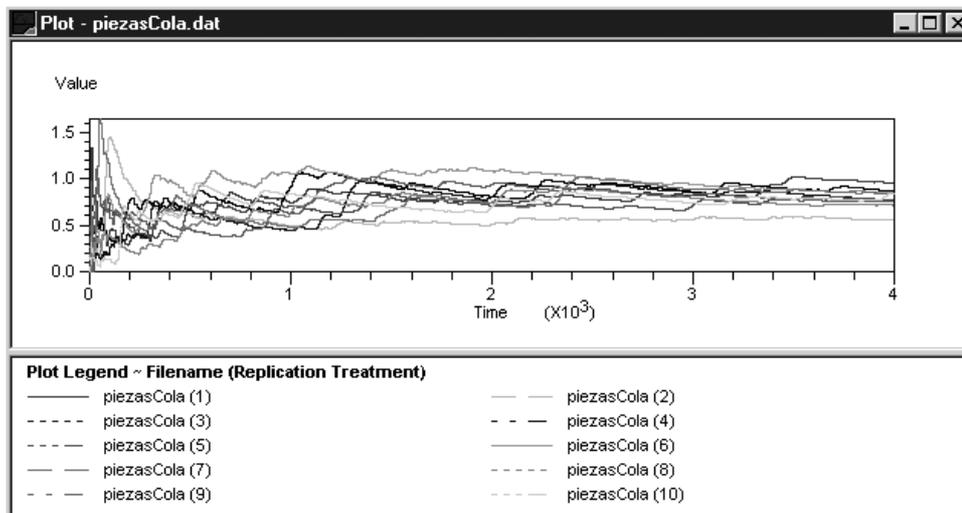


Figura 11.17: Evolución del número medio de piezas en cola.

Para analizar los datos usando otras herramientas software es conveniente exportar los datos a un fichero de texto. Output Analyzer puede convertir el fichero “.dat” en un fichero de texto ascii. Para ello debe seleccionarse “File / Data File / Export”, y especificar el nombre (casilla “Export to File”) del fichero al cual se desean volcar los datos (por ejemplo, piezasCola.txt). □

Una vez identificada la duración del estacionario, puede emplearse el método de réplicas/eliminación, o bien el método de la media por lotes, para estudiar el comportamiento del sistema en el estacionario. En ambos casos, debe indicarse la duración del periodo transitorio (“Warm-up Period”) al definir el experimento.

Si la simulación del periodo transitorio no requiere un tiempo muy elevado, el análisis en el estacionario puede hacerse mediante el método de réplicas/eliminación. Se realizan réplicas independientes de la simulación, al igual que se hace en el análisis con terminación, con la diferencia de que en este caso la longitud del periodo transitorio es distinta de cero. Arena construye el intervalo del 95% de confianza a partir de las n observaciones obtenidas en las n réplicas independientes de la simulación (una observación por réplica).

Esta idea se aplica igualmente cuando se comparan varias alternativas mediante PAN. Sin embargo, hay que tener en cuenta que la longitud del transitorio puede variar de un escenario a otro. Es preciso estudiar la longitud del transitorio de cada uno de ellos y escoger aquel que sea mayor.

En aquellos casos en que la simulación del periodo transitorio es costosa, el método de réplicas/eliminación no resulta eficiente, ya que en todas las réplicas se simula el transitorio. En estos casos, es preferible aplicar el método de la media por lotes, en el que se simula el transitorio tan sólo una vez. El procedimiento se basa en obtener observaciones no correlacionadas entre sí, dividiendo la secuencia de observaciones de la respuesta en subsecuencias (o lotes), y calculando la media de las observaciones de cada uno de ellos. Cada una de estas medias juega el mismo papel que juega una observación, obtenida de una réplica independiente, en el método de réplicas/eliminación.

Cuando el número de réplicas de la simulación es uno, Arena aplica el método de la media por lotes: construye el intervalo del 95% de confianza a partir de las medias obtenidas de las subsecuencias. Es importante que la longitud de las subsecuencias sea grande, a fin de que los valores medios obtenidos (uno por subsecuencia) no estén correlacionados

entre sí. Arena¹² trata de escoger las subsecuencias lo suficientemente largas como para que aparentemente no estén correlacionadas, y si la longitud de la simulación es demasiado corta para ello, lo indica no calculando la longitud del semi-intervalo. Los datos obtenidos durante el periodo transitorio no se consideran en el análisis, es decir, Arena comienza la primera subsecuencia una vez finalizado el transitorio.

Es posible controlar la selección de las subsecuencias. Para ello deben volcarse los datos de la simulación desde Arena a un fichero “.dat”, cargarlos desde Output Analyzer, y usar la opción “Analyze / Batch Truncate” para definir las subsecuencias y calcular la media de cada una de ellas. Al realizar la división en subsecuencias, Output Analyzer realiza los mismos test de correlación que Arena, e indica si detecta correlación. Algunas razones para querer usar Output Analyzer en este caso son:

- Cuando se desea construir un intervalo con una confianza diferente del 95 %.
- Cuando se quieren comparar dos alternativas, basándose en su comportamiento en el estacionario.

EJERCICIOS

11.1 Realice el estudio de simulación descrito a continuación¹³, usando para ello el entorno de modelado Arena y la herramienta *Arena Output Analyzer*.

La estructura lógica del sistema bajo estudio y sus aspectos numéricos son lo siguientes. A un sistema compuesto por dos máquinas llegan piezas para ser procesadas. El intervalo de tiempo entre llegadas está distribuido exponencialmente, con media 20 minutos. Al llegar, las piezas son enviadas a la Máquina 1 y procesadas. El tiempo de proceso sigue una distribución TRIA(4.5, 9.3, 11) minutos. A continuación, las máquinas son procesadas en la Máquina 2, donde el tiempo de proceso está distribuido TRIA(16.4, 19.1, 21.8). Las piezas son enviadas de nuevo a la Máquina 1, para ser procesadas una segunda vez (con el mismo tiempo de proceso). Seguidamente, las piezas abandonan el sistema.

Condiciones iniciales: no hay ninguna pieza en el sistema. Condición de finalización: simular el funcionamiento del sistema ininterrumpidamente durante 20000 minutos.

El estudio tiene un doble objetivo. En primer lugar, se desean estimar los dos siguientes estadísticos:

- El número promedio de piezas en la cola de cada máquina.
- El tiempo de ciclo medio de las piezas.

para el sistema tal como se ha descrito anteriormente, realizando para ello una única réplica de la simulación.

En segundo lugar, se desea comparar estos valores con los que se obtienen si el tiempo de proceso en el segundo paso por la Máquina 1 estuviera distribuido TRIA(6.7, 9.1, 13.6). Para realizar la comparación, deben hacerse 20 réplicas de cada una de las versiones del modelo, y calcular el intervalo del 95 % de confianza de la diferencia entre las respuestas, usando Output Analyzer.

11.2 A una delegación de Tráfico llegan dos tipos de clientes¹⁴:

- Clientes interesados en temas relacionados con los vehículos (altas, bajas, matriculación, etc.). El intervalo de tiempo entre llegadas consecutivas de estos clientes está distribuido EXPO (6.8), y los tiempos de servicio TRIA (8.7, 13.7, 15.2). Todos los tiempos están expresados en minutos.
- Clientes interesados en temas relacionados con su carnet de conducir. El tiempo entre llegadas está distribuido EXPO (8.7), y el tiempo de servicio TRIA (16.7, 20.5, 29.2).

Se forman dos colas: una para cada tipo de cliente. Hay cinco trabajadores: dos dedicados a los vehículos, otros dos dedicados a los conductores, y un supervisor, que puede atender a ambos tipos de clientes.

¹²En el texto (Kelton et al. 2002) se explica cómo realiza Arena la selección de las subsecuencias.

¹³Este estudio de simulación está extraído del texto (Kelton et al. 2002).

¹⁴Este estudio de simulación está extraído del texto (Kelton et al. 2002).

- a) Suponga que los 5 empleados trabajan 8 horas al día. Realice 30 réplicas de la simulación, y calcule los intervalos del 95% de confianza del tiempo de ciclo medio de cada tipo de cliente.
 - b) Analice el sistema de nuevo, suponiendo que uno de los trabajadores que atiende los temas de los vehículos está también capacitado para atender los temas relacionados con los conductores. Realice 30 réplicas y estime el valor esperado de la diferencia entre esta configuración y la anterior (basándose en el tiempo de ciclo medio).
- 11.3** Estime el número medio de piezas en cola, en el estado estacionario, del proceso de perforación descrito en el Problema 6.1.

Tema 12

Técnicas de reducción de la varianza

NOTA IMPORTANTE: Este tema NO SE EXIGIRÁ EN EL EXAMEN durante el presente curso académico, si bien se recomienda su lectura. Por este motivo, no se plantean objetivos docentes del tema, ni tampoco se proponen problemas.

12.1. Introducción

En algunos estudios, el número de muestras experimentales que deben obtenerse mediante simulación, para poder extraer conclusiones con el nivel de confianza deseado, es muy alto. Cuando este número es irrealizable, los resultados del estudio se obtienen con una precisión pobre: los intervalos de confianza del valor esperado de las magnitudes de interés son demasiado anchos.

Existen varios factores que influyen en la anchura del intervalo de confianza. Supóngase que se realizan n observaciones de una variable aleatoria X y que se construye un intervalo de confianza para la media, μ_X . Considérense los dos intervalos siguientes:

$$\bar{X}(n) \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \quad (12.1)$$

$$\bar{X}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \quad (12.2)$$

En ambos casos, la anchura del intervalo depende de:

- El tanto por ciento de confianza del intervalo, $100 \cdot (1 - \alpha) \%$, afecta a la elección del punto crítico de la distribución. Al aumentar el tanto por ciento de confianza, aumenta la anchura del intervalo.
- El tamaño n de la muestra usada para la construcción del intervalo. La anchura del intervalo es proporcional a $\frac{1}{\sqrt{n}}$. En consecuencia, aumentando n disminuye la anchura del intervalo.

- El estimador de la varianza de la muestra usada para la construcción del intervalo, $S_X^2(n)$. La anchura del intervalo es proporcional a $\sqrt{S_X^2(n)}$, con lo cual, disminuyendo $S_X^2(n)$ disminuye la anchura del intervalo de confianza.

Este último punto tiene una gran importancia práctica. Si de alguna forma se consigue reducir la varianza de la variable aleatoria X , sin modificar su valor esperado μ_X , entonces se puede obtener mayor precisión en la estimación del valor esperado (menor anchura del intervalo de confianza) con el mismo número n de observaciones de X . Las técnicas desarrolladas para perseguir este objetivo se denominan *técnicas de reducción de la varianza*.

Algunas características generales de estas técnicas son las siguientes:

- La forma de aplicar la técnica a un problema en concreto depende del modelo en particular. Por tanto, su aplicación requiere un conocimiento profundo del funcionamiento del modelo.
- Generalmente es imposible cuantificar de antemano qué reducción en la varianza se conseguirá al aplicar la técnica. Peor aun, tampoco hay garantía de que realmente se consiga reducir la varianza. La forma de despejar estas incógnitas es realizar algunas réplicas piloto de la simulación y comparar los resultados de aplicar técnicas de reducción de la varianza con los de realizar directamente la simulación sin aplicarlas.
- Algunas técnicas de reducción de la varianza aumentan la carga computacional, con lo cual es preciso sopesar si la carga computacional ahorrada de la reducción de la varianza es mayor o menor que la carga computacional de aplicar la técnica.

En esta sección se describen únicamente dos de las muchas¹ técnicas de reducción de la varianza existentes:

- la técnica de los *números aleatorios comunes*, aplicable cuando se comparan dos o más sistemas, y
- la técnica de las *variables antitéticas*, aplicable al estudio de un único sistema.

12.2. Números pseudo aleatorios comunes

La técnica se basa en la idea de comparar los sistemas *bajo condiciones experimentales similares*, de modo que aumente la seguridad de que las diferencias de comportamiento observadas son debidas a la naturaleza de los sistemas, y no a fluctuaciones en las condiciones experimentales.

En este contexto, se entiende por *condiciones experimentales* las observaciones de las variables aleatorias generadas para alimentar el modelo durante la simulación. Por ejemplo, en las simulaciones de colas éstas pueden ser: los intervalos de tiempo entre llegadas, las necesidades de servicio de los clientes, etc.

El nombre de la técnica proviene de que en bastantes ocasiones es posible emplear la misma secuencia de números aleatorios para alimentar los dos sistemas que se pretende comparar.

Lógica del método

Se parte del supuesto de que el objetivo del estudio es comparar el valor esperado de la variable X_1 , del sistema 1, con el valor esperado de la variable X_2 , del sistema 2. Para ello, debe aplicarse el procedimiento descrito en la Sección 11.5:

1. Se realizan n réplicas de la simulación de cada sistema, obteniéndose n observaciones de X_1 y n observaciones de X_2 .

¹En (Law & Kelton 2000) se proporcionan algunas referencias bibliográficas en las que se realiza una clasificación de las técnicas de reducción de la varianza aplicables a simulación. Estos artículos contienen a su vez otras muchas referencias en las que se describen las diferentes técnicas.

2. Se emparejan las observaciones de X_1 y X_2 , y se realiza la resta, obteniéndose n observaciones de la variable aleatoria $Z = X_1 - X_2$.

$$z_1 = x_{1,1} - x_{2,1}, \quad \dots \quad z_n = x_{1,n} - x_{2,n} \quad (12.3)$$

3. A partir de las n observaciones de Z , calculadas en el paso anterior, se construye un intervalo de confianza para el valor esperado de Z , $\mu_Z = \mu_{X_1} - \mu_{X_2}$:

$$\bar{Z}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_Z^2(n)}{n}} \quad (12.4)$$

La anchura del intervalo de confianza es proporcional a $\frac{S_Z^2(n)}{n}$, que es el estimador de la varianza del estimador $\bar{Z}(n)$. Así pues, cuanto menor sea la varianza de $\bar{Z}(n)$, menor será la anchura del intervalo. El objetivo del método es obtener observaciones de X_1 y X_2 , tales que la varianza del estimador de la media de Z sea la menor posible. Para ello, se saca partido de la propiedad siguiente:

$$Z = X_1 - X_2 \rightarrow \sigma_Z^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 - 2 \cdot \text{cov}(X_1, X_2) \quad (12.5)$$

La interpretación de la *covarianza* es muy intuitiva. La *covarianza* de X_1 y X_2 es una medida de la dependencia lineal entre ambas variables aleatorias:

$$\text{cov}(X_1, X_2) = E\{(X_1 - \mu_{X_1}) \cdot (X_2 - \mu_{X_2})\} \quad (12.6)$$

- Cuando $\text{cov}(X_1, X_2) > 0$, significa que tiende a ocurrir uno de los siguientes dos eventos o ambos:

$$\{X_1 > \mu_{X_1}, X_2 > \mu_{X_2}\} \quad \{X_1 < \mu_{X_1}, X_2 < \mu_{X_2}\} \quad (12.7)$$

Es decir, cuando una de las dos variables toma un valor grande, la otra también tiende a ser grande. Por el contrario, cuando una de las variables toma un valor pequeño, la otra tiende a ser pequeña. Se dice que las variables están *correlacionadas positivamente*.

- Cuando $\text{cov}(X_1, X_2) < 0$, significa que tiende a ocurrir uno de los siguientes dos eventos o ambos:

$$\{X_1 > \mu_{X_1}, X_2 < \mu_{X_2}\} \quad \{X_1 < \mu_{X_1}, X_2 > \mu_{X_2}\} \quad (12.8)$$

Es decir, cuando una variable toma un valor grande, la otra tiende a tomar un valor pequeño, y viceversa. Se dice que ambas variables están *correlacionadas negativamente*.

- Cuando $\text{cov}(X_1, X_2) = 0$, se dice que las dos variables *no están correlacionadas*. Si dos variables son independientes, entonces no están correlacionadas. Sin embargo, el hecho de que dos variables no estén correlacionadas no implica que sean independientes.

De la discusión anterior se deduce que si la simulación de los sistemas, 1 y 2, se realiza empleando secuencias independientes de números pseudo aleatorios, entonces X_1 y X_2 serán independientes. Con lo cual:

$$\text{cov}(X_1, X_2) = 0 \rightarrow \sigma_Z^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 \quad (12.9)$$

Por otro lado, si X_1 y X_2 están positivamente correlacionadas, entonces:

$$\sigma_Z^2 < \sigma_{X_1}^2 + \sigma_{X_2}^2 \quad (12.10)$$

con lo que se reduce la varianza del estimador $\bar{Z}(n)$. Esta reducción de la varianza del estimador hace que aumente la probabilidad de que el valor del estimador obtenido de las observaciones se encuentre próximo a μ_Z . En consecuencia se acorta la longitud del intervalo de confianza para el valor esperado de Z .

Aplicabilidad de la técnica

La técnica de los números seudo aleatorios comunes pretende inducir correlación positiva entre las variables aleatorias X_1 y X_2 . Para ello, se usa la misma secuencia de números seudo aleatorios para simular ambos sistemas. La eficacia de esta técnica depende del hecho de que las variables de interés del modelo estén correlacionadas positivamente cuando se excitan los modelos de la misma forma.

Por el contrario, si excitando ambos sistemas de la misma forma, las variables de interés reaccionan en direcciones opuestas (están correlacionadas negativamente), entonces el empleo de la técnica es contraproducente: produce un aumento de la varianza. A continuación se muestra un ejemplo de correlación positiva.

Ejemplo 12.2.1. *Cabe esperar que, en (casi) cualquier sistema, intervalos de tiempo mayores entre llegadas producirán tiempos de espera menores en las colas. Igualmente, intervalos entre llegadas menores producirán tiempos de espera mayores en las colas. Por tanto, las variables “tiempo de espera en cola” de diferentes sistemas estarán correlacionadas positivamente entre sí, cuando se usen los mismos números seudo aleatorios para generar los intervalos entre llegadas. La conclusión de ello es que existe cierto tipo de sistemas para los cuales el éxito en la aplicación de esta técnica está garantizado.* \square

A la hora de realizar razonamientos del tipo “las variables reaccionan en la misma dirección, o en direcciones opuestas, ante números seudo aleatorios iguales”, debe tenerse en cuenta qué procedimiento se sigue para generar las observaciones de las variables aleatorias con que se alimenta el modelo. Estas observaciones se calculan a partir de números seudo aleatorios siguiendo un determinado método, por ejemplo, el de la transformación inversa. Una propiedad deseable de estos métodos es que exista una relación monótona creciente entre el número seudo aleatorio, y el valor de la variable aleatoria generada a partir de él. El método de la transformación inversa tiene esta propiedad.

Así, por ejemplo, si los intervalos entre llegadas están distribuidos exponencialmente y se usa el método de la transformación inversa, entonces números seudo aleatorios grandes producirán intervalos entre llegadas grandes, y números seudo aleatorios pequeños producirán intervalos pequeños. A partir de esto, podrá suponerse que números seudo aleatorios grandes (al generar intervalos entre llegadas grandes), producirán tiempos pequeños de espera en cola.

Con el fin de decidir si resulta ventajoso aplicar en un determinado estudio la técnica de los números seudo aleatorios comunes, puede realizarse el estudio piloto siguiente:

1. Realizar n réplicas de la simulación de cada sistema, usando en ambos casos la misma secuencia de números seudo aleatorios. Se obtienen las observaciones:

$$\begin{aligned} x_{1,1}, \dots, x_{1,n} & \quad (n \text{ observaciones de } X_1) \\ x_{2,1}, \dots, x_{2,n} & \quad (n \text{ observaciones de } X_2) \end{aligned} \quad (12.11)$$

Un procedimiento para hacerse una idea de hasta qué punto se ha conseguido que X_1 y X_2 estén correlacionadas positivamente, consiste en representar gráficamente las observaciones de X_2 frente a las observaciones de X_1 , es decir, los puntos $(x_{1,j}, x_{2,j})$ para $j : 1, \dots, n$. En la Figura 12.1 se muestra un ejemplo.

2. Emparejando y restando las observaciones anteriores, calcular las n muestras de Z . A continuación, a partir de las n muestras de X_1 , X_2 y Z , calcular el estimador de la varianza de cada una de estas tres variables aleatorias: $S_{X_1}^2(n)$, $S_{X_2}^2(n)$, y $S_Z^2(n)$.

Si se hubieran utilizado para un sistema números seudo aleatorios independientes de los usados para el otro, entonces X_1 y X_2 serían independientes y $S_{X_1}^2(n) + S_{X_2}^2(n)$ sería un estimador de σ_Z^2 . Por tanto, si:

$$S_Z^2(n) < S_{X_1}^2(n) + S_{X_2}^2(n) \quad (12.12)$$

es indicativo de que el método está dando resultados. La diferencia entre $S_{X_1}^2(n) + S_{X_2}^2(n)$ y $S_Z^2(n)$ puede considerarse una estimación de su eficacia.

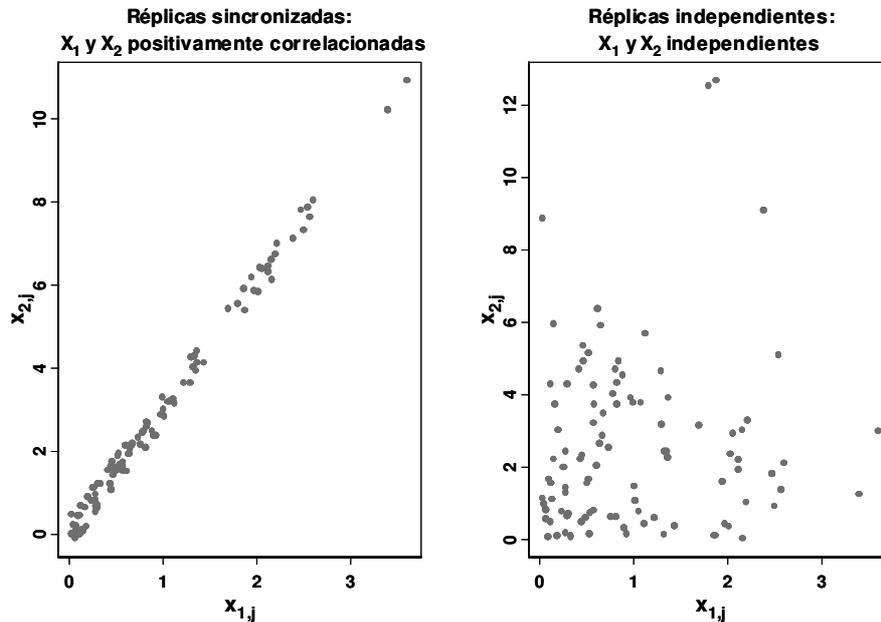


Figura 12.1: Correlación positiva (izqda). No correlación (drcha).

Sincronización

Para aplicar la técnica es preciso:

- Emplear en la réplica i de la simulación, las mismas secuencias de números pseudo aleatorios para los dos sistemas.
- En cada réplica debe usarse una secuencia independiente de las empleadas en las demás réplicas.

Idealmente, un determinado número pseudo aleatorio, empleado para un propósito específico en la simulación de un sistema, debe ser empleado para exactamente el mismo propósito en la simulación del otro sistema. Por ejemplo, si se emplea el número pseudo aleatorio, u_i , para generar un determinado tiempo de servicio en la simulación del sistema 1, entonces debe usarse en la simulación del sistema 2 para generar una observación del mismo tiempo de servicio.

Como pretende ilustrar el siguiente ejemplo, esto no se consigue con únicamente inicializar los generadores de números aleatorios con el mismo valor de la semilla. No basta con “reutilizar” la misma secuencia de números pseudo aleatorios, sin prestar atención a cómo se usa cada uno.

Ejemplo 12.2.2. *Considérese de nuevo el estudio de simulación planteado en el Ejemplo 11.5.1. Se pretende comparar el funcionamiento del sistema de venta de billetes, cuando se utiliza una máquina de tipo ZZ, con el funcionamiento cuando se usan dos máquinas de tipo Z.*

La medida de interés en ambos casos es el tiempo medio de espera por cliente, calculado a partir del tiempo de espera de los primeros 100 clientes que inician el proceso de compra del billete. La condición inicial en ambos sistemas es que el sistema está libre, y la cola vacía cuando llega el primer cliente.

En un intento (mal planteado) de aplicar la técnica de los números pseudo aleatorios comunes, se usa la misma secuencia de números aleatorios para generar los tiempos entre llegadas y los tiempos de servicio de ambos sistemas. Para ello, simplemente se usa la misma semilla para inicializar la réplica i de la simulación de ambos sistemas (se usa una semilla diferente para

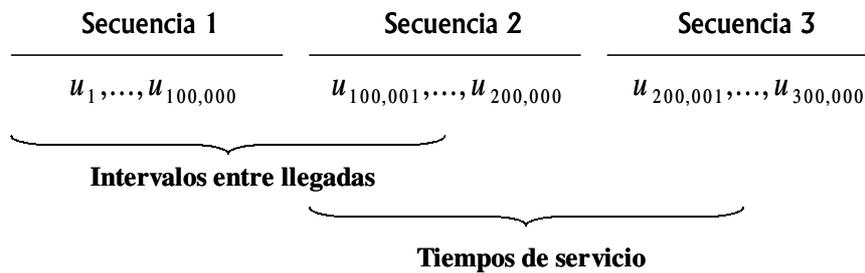


Figura 12.2: Solapamiento de las secuencias de números pseudo aleatorios.

diferentes réplicas). El ejemplo pretende demostrar que, en general, esto no es suficiente para garantizar la correcta aplicación de la técnica de reducción de la varianza.

En la primera réplica de la simulación del sistema 1, se ha necesitado una secuencia de 219 números pseudo aleatorios, mientras que en la primera réplica de la simulación del sistema 2 sólo se han necesitado 201 números pseudo aleatorios. La cantidad de números pseudo aleatorios que se necesita en una determinada réplica de la simulación es una variable aleatoria. En consecuencia, cualquiera de los dos sistemas puede requerir más números pseudo aleatorios que el otro.

En la Tabla 12.1 se muestran algunos valores de la secuencia de números pseudo aleatorios usada en esta réplica, y se indica el uso que se le ha dado al simular cada uno de los sistemas. Se representa “LL” si el número se ha empleado como intervalo entre llegadas, y “S” si se ha usado como tiempo de servicio.

Como se observa, el uso general (LL vs S) de los primeros 8 números pseudo aleatorios coincide en ambos sistemas². Sin embargo, u_9 se usa en el sistema ZZ como intervalo entre llegadas y en el sistema con dos máquinas Z como tiempo de servicio. A partir de ese punto, se mezcla el uso de los números pseudo aleatorios: existe sincronización aproximadamente en el 50% de los usos. □

Existen trucos para mantener la sincronización, que se pueden emplear en la programación del modelo. Algunos de ellos son los siguientes:

- Emplear múltiples secuencias de números aleatorios, de modo que cada secuencia se dedica a producir observaciones de un determinado tipo de variables aleatorias de entrada. En la simulación de una cola, por ejemplo, se puede dedicar una secuencia para generar intervalos entre llegadas, y una secuencia diferente para generar tiempos de servicio.

La dedicación de las secuencias de números aleatorios es generalmente una buena idea, y la mayoría de los lenguajes de simulación soportan la capacidad de separar las secuencias. Sin embargo, dado que las secuencias de números aleatorios suelen ser segmentos adyacentes de la salida de un determinado generador de números aleatorios, y por tanto tienen una determinada longitud (frecuentemente 100000 números pseudo aleatorios por secuencia), debe evitarse que las secuencias se solapen cuando se realizan simulaciones largas, o cuando se realizan muchas réplicas. Para ello, se puede realizar una estimación aproximada de cuantos números pseudo aleatorios van a ser necesarios en el estudio de simulación. A continuación se muestra un ejemplo.

Ejemplo 12.2.3. Si se va a realizar la simulación del sistema ZZ y se espera que pasen aproximadamente 5000 clientes por réplica, hacen falta 5000 intervalos entre llegadas y 5000 tiempos de servicio. Si se usa el método de la transformada inversa para generar las observaciones de estas variables aleatorias de entrada, hacen falta 5000 números pseudo aleatorios de cada secuencia.

²Este ejemplo está extraído del texto (Law & Kelton 2000).

Tabla 12.1: Uso de los números pseudo aleatorios en una réplica no sincronizada.

| j | u_j | Uso en ZZ | Uso en Z | ¿Concuerda? |
|-----|-------|-----------|----------|-------------|
| 1 | 0.40 | LL | LL | SI |
| 2 | 0.61 | LL | LL | SI |
| 3 | 0.43 | S | S | SI |
| 4 | 0.38 | LL | LL | SI |
| 5 | 0.51 | S | S | SI |
| 6 | 0.71 | LL | LL | SI |
| 7 | 0.19 | S | S | SI |
| 8 | 0.83 | LL | LL | SI |
| 9 | 0.65 | LL | S | NO |
| 10 | 0.38 | LL | LL | SI |
| 11 | 0.35 | S | S | SI |
| 12 | 0.76 | LL | LL | SI |
| 13 | 0.45 | LL | LL | SI |
| 14 | 0.91 | S | LL | NO |
| 15 | 0.47 | S | LL | NO |
| 16 | 0.48 | LL | S | NO |
| 17 | 0.85 | S | S | SI |
| 18 | 0.80 | S | S | SI |
| 19 | 0.72 | S | LL | NO |
| 20 | 0.70 | LL | LL | SI |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 195 | 0.60 | LL | S | NO |
| 196 | 0.42 | LL | S | NO |
| 197 | 0.57 | LL | S | NO |
| 198 | 0.86 | LL | LL | SI |
| 199 | 0.10 | LL | S | NO |
| 200 | 0.15 | S | LL | NO |
| 201 | 0.33 | S | S | SI |
| 202 | 0.58 | LL | | |
| 203 | 0.83 | LL | | |
| ⋮ | ⋮ | ⋮ | | |
| 218 | 0.34 | LL | | |
| 219 | 0.66 | S | | |

Si se realizan 30 réplicas de la simulación, hacen falta 150000 números pseudo aleatorios de cada secuencia. Si las secuencias tienen una longitud de 100000 números y dedicamos la secuencia 1 a los intervalos entre llegadas y la secuencia 2 a los tiempos de servicio, entonces los últimos 50000 números de los intervalos entre llegadas son los primeros 50000 números usados para los tiempos de servicio (ver Figura 12.2), lo cual hace que las réplicas no sean independientes entre sí.

La solución es “saltarse” secuencias: usar la secuencia 1 (y la mitad de la secuencia 2) para los tiempos entre llegadas y usar, por ejemplo, la secuencia 6 (y la mitad de la secuencia 7) para los tiempos de servicio. \square

- El empleo del método de la transformación inversa, para generar las observaciones de las variables aleatorias de entrada, facilita la sincronización, ya que siempre se necesita precisamente un número pseudo aleatorio para generar una observación.

A pesar de las recomendaciones anteriores, en ocasiones no es posible sincronizar la simulación de todos los modelos bajo estudio. Incluso, el trabajo extra que ello supone puede no verse compensado por la reducción en la varianza obtenida. No debe descartarse la posibilidad, cuando se juzgue que es lo más conveniente, de sincronizar algunas de las variables de entrada y generar el resto independientemente. Por ejemplo, en un modelo con una compleja red de colas, puede optarse por sincronizar los intervalos entre llegadas pero no los tiempos de servicio.

12.3. Variables aleatorias antitéticas

Esta técnica de reducción de la varianza es aplicable a la simulación de un único sistema: el objetivo del estudio es construir un intervalo de confianza para el valor esperado, μ_X , de cierta magnitud aleatoria de interés, X .

La técnica pretende reducir la varianza de las muestras x_1, \dots, x_n obtenidas (una por réplica) y, en consecuencia, reducir la anchura del intervalo de confianza. Para ello, se intenta inducir correlación negativa entre las replicas de la simulación del sistema de la manera siguiente.

Lógica del método

Se realizan n parejas de réplicas de la simulación del sistema, obteniéndose una observación de cada réplica (en total $2 \cdot n$ observaciones de X). Se representan de la forma siguiente:

$$\begin{array}{ccc} x_1^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & \dots & x_n^{(2)} \end{array}$$

donde $x_j^{(1)}$ y $x_j^{(2)}$ son las observaciones obtenidas en la pareja j de réplicas.

La técnica consiste en conseguir que *si una de las observaciones de la pareja es grande, entonces la otra tienda a ser pequeña, y viceversa*. Es decir, se trata de que las dos observaciones de la pareja de réplicas estén *correlacionadas negativamente*.

A continuación, el intervalo de confianza para μ_X se construye usando, como dato básico de análisis, el promedio de las observaciones de la pareja:

$$x_j = \frac{x_j^{(1)} + x_j^{(2)}}{2} \quad \text{para } j : 1, \dots, n \quad (12.13)$$

Cabe suponer que estos promedios (el punto central entre una observación grande y una pequeña) estarán más próximos al valor esperado, μ_X , que si las observaciones se hubieran considerado independientemente. En consecuencia, se reducirá la varianza del estimador de μ_X .

Esta idea intuitiva puede formalizarse, teniendo en cuenta que:

$$\text{var} \{ \bar{X}(n) \} = \frac{\text{var} \{ X_j \}}{n} = \frac{\text{var} \{ X_j^{(1)} \} + \text{var} \{ X_j^{(2)} \} + 2 \cdot \text{cov} \{ X_j^{(1)}, X_j^{(2)} \}}{4 \cdot n} \quad (12.14)$$

Si las réplicas se realizaran independientemente, entonces:

$$\text{cov} \{ X_j^{(1)}, X_j^{(2)} \} = 0 \quad (12.15)$$

Sin embargo, si se consigue que:

$$\text{cov} \{ X_j^{(1)}, X_j^{(2)} \} < 0 \quad (12.16)$$

entonces se reduce $\text{var} \{ \bar{X}(n) \}$ (la varianza del estimador de μ_X), y en consecuencia se reduce la anchura del intervalo de confianza.

Aplicabilidad de la técnica

Como ocurre con las demás técnicas de reducción de la varianza, no es posible saber de antemano el beneficio que se conseguirá al aplicarla en un estudio de simulación en concreto. Para estimarlo puede realizarse un estudio piloto, como el descrito en la sección anterior.

Un requisito fundamental para que la técnica consiga reducir la varianza es que la respuesta del sistema sea una función monótona (en cualquier dirección) de los números pseudo aleatorios. Hay determinadas magnitudes que frecuentemente satisfacen este requisito. Por ejemplo, en la mayoría de los modelos de colas, al disminuir el intervalo entre llegadas, y al aumentar el tiempo de servicio, se produce una congestión de la cola: aumenta monótonamente el tiempo medio de espera. El ingeniero debe emplear su conocimiento del funcionamiento del modelo, para decidir si cabe esperar una relación monótona (creciente o decreciente) entre los números pseudo aleatorios de entrada y la variable de interés.

Sincronización

La forma más sencilla de introducir correlación negativa entre una pareja de réplicas es usar números pseudo aleatorios complementarios. Es decir, si se usa el número pseudo aleatorio u_k en una de las réplicas para un propósito particular, entonces emplear $1 - u_k$ en la otra réplica para ese mismo propósito. Obsérvese que esto no supone ningún problema, ya que:

$$U \sim U(0, 1) \quad \leftrightarrow \quad 1 - U \sim U(0, 1) \quad (12.17)$$

Por ejemplo, si en una réplica se usa un número pseudo aleatorio grande (cercano a 1) para generar un tiempo de servicio, entonces el tiempo de servicio será grande³: se congestiona la cola. El número pseudo aleatorio complementario, $1 - U$, será pequeño. El efecto de este tiempo de servicio pequeño en la segunda réplica es el deseado: descongestionar la cola.

La técnica se aplica con éxito porque se usan para el mismo propósito el número en la primera réplica y su simétrico en la segunda réplica. Sin embargo, ¿qué ocurriría si se usara el número en la primera réplica para generar un tiempo de servicio y el número simétrico en la segunda réplica para generar un intervalo entre llegadas? El tiempo de servicio grande

³Se supone que se usa un método como el de la transformación inversa, que hace corresponder un número pseudo aleatorio grande con un valor grande de la variable aleatoria:

Si u_k es grande, entonces $F^{-1}(u_k)$ es grande y $F^{-1}(1 - u_k)$ es pequeño, ya que la función inversa de la probabilidad acumulada, F^{-1} , es monótona creciente.

congestiona la cola en la primera réplica, el mismo efecto que produce el intervalo entre llegadas pequeño en la segunda réplica. Se está introduciendo correlación positiva, justo el efecto contrario al deseado.

Por ello, en la aplicación de la técnica es recomendable usar secuencias de números aleatorios diferentes para generar variables aleatorias de tipo diferente.

Tema 13

Diseño de experimentos y optimización

Objetivos: Una vez estudiado el contenido del tema y realizados los ejercicios prácticos, debería saber:

- Discutir los conceptos esenciales del diseño de experimentos.
- Diseñar experimentos con uno o varios factores.
- Interpretar los resultados de experimentos para analizar varios factores simultáneamente.
- Realizar experimentos usando Arena.

13.1. Introducción

En este tema se realiza una introducción al diseño estadístico de experimentos y a las técnicas de optimización, aplicando estas metodologías al caso en el cual el experimento es la simulación de un modelo mediante el ordenador.

El *diseño de experimentos* constituye una aproximación estructurada al estudio del impacto que tienen los parámetros del modelo sobre su respuesta. Este tipo de estudios se realiza típicamente cuando:

- Se desean estudiar relaciones causa-efecto entre determinados parámetros del modelo (a los que se denomina *factores*) y determinadas medidas del comportamiento del modelo (a las que se denomina *respuestas*).
- Se desea investigar qué factores tienen un mayor impacto sobre la respuesta.

Un objetivo de estudio íntimamente relacionado con los anteriores es la *optimización*. Se trata de estimar qué valores de los factores maximizan (o minimizan) la respuesta. Típicamente los estudios de optimización se realizan en dos pasos:

- En el primer paso, se estudia en qué medida afecta la variación en el valor de los factores a la respuesta. El objetivo de ello es determinar qué factores tienen un impacto significativo sobre la respuesta, y cuáles tienen un impacto despreciable. Este primer estudio se realiza usando las técnicas de *diseño de experimentos*.

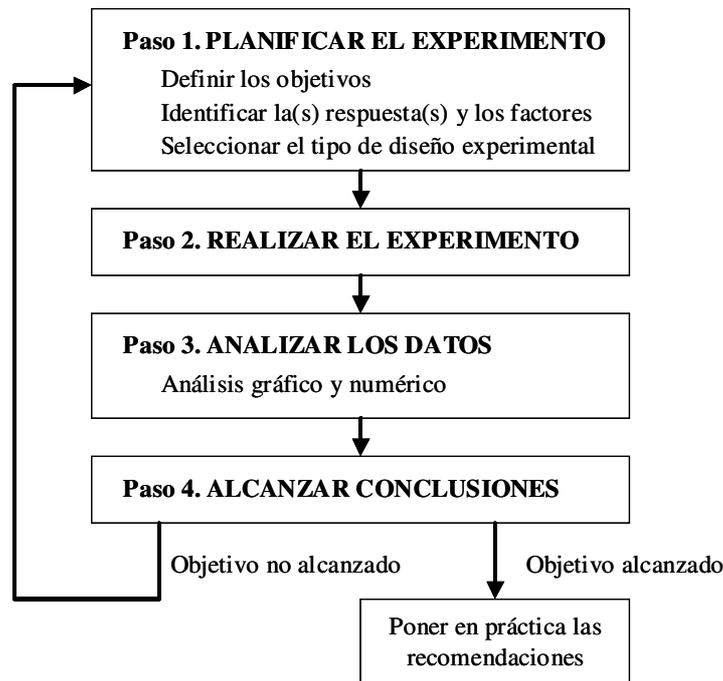


Figura 13.1: Pasos en el diseño de experimentos.

- Una vez identificadas qué combinaciones de factores tienen un efecto de primer orden (es decir, significativo) sobre la respuesta, se plantea el problema de la *optimización*: encontrar para qué valores de estos factores se optimiza el valor de la respuesta.

La razón por la cual es recomendable planificar los estudios de simulación, usando técnicas de diseño de experimentos, es la economía de medios. La necesidad de realizar estos estudios de forma estructurada surge cuando se desea estudiar la influencia simultánea de varios factores.

Cuando únicamente se desea estudiar un factor, la planificación es sencilla: basta con simular el modelo para determinados valores del factor. Estos valores del factor se denominan *niveles*, en la terminología del diseño de experimentos. Realizando para cada nivel del factor un determinado número de réplicas de la simulación, puede construirse para cada nivel un intervalo de confianza para la respuesta.

Sin embargo, cuando se desea estudiar la influencia simultánea sobre la respuesta de $k > 1$ factores, el objetivo del estudio es más amplio:

- se desea estimar cómo influye cada factor individualmente sobre la respuesta, y
- qué *interacciones* existen entre los factores, es decir, en qué medida el efecto de un factor sobre la respuesta depende de los niveles de los restantes factores.

Un modo (no recomendable) de plantear este estudio es estimar el efecto de cada factor variando su valor, a la vez que se mantienen fijos los valores de los $k - 1$ factores restantes. Se repite el proceso k veces, a fin de examinar cada uno de los factores a un tiempo. Esta estrategia es poco eficiente, considerando el número de combinaciones niveles que hay que simular, con relación a la información obtenida. En concreto, esta estrategia no permite estimar las *interacciones* entre los factores.

El diseño de experimentos proporciona un modo de decidir qué combinaciones de niveles deben ser estudiadas, a fin de obtener de la manera más eficiente la información requerida.

Se trata, por tanto, de una metodología cuya finalidad es *racionalizar y optimizar el esfuerzo experimental*.

En la Figura 13.1 se muestran los cuatro pasos básicos en el diseño de experimentos. La planificación del experimento es un proceso iterativo que comprende las siguientes tareas:

1. Definir los objetivos del experimento.
2. Seleccionar las respuestas, en consonancia con los objetivos del experimento.
3. Seleccionar aquellos factores que previsiblemente influyen en el valor de las respuestas.
4. Seleccionar el nivel de los factores.
5. Seleccionar el diseño experimental. Existen varios tipos de diseños experimentales: factorial completo, factorial fraccional, diseños robustos, la metodología de la superficie de respuesta (central compuesto y Box-Behnken), etc. La selección de un tipo de diseño u otro depende de principalmente de:
 - Los objetivos del estudio.
 - El número de simulaciones que puedan razonablemente realizarse.
 - Las respuestas, los factores y el número de niveles de cada factor.
 - El tipo de modelo matemático que se desee ajustar a los datos. Cada diseño experimental lleva implícito un modelo matemático, que relaciona la respuesta con los factores.

13.2. Objetivos, respuestas y factores

Objetivos del experimento

A la hora de fijar los objetivos de un experimento, debe tenerse en cuenta que generalmente es preferible realizar una serie de pequeños experimentos, de manera secuencial, a realizar un único experimento diseñado para responder todas las posibles cuestiones.

La realización de experimentos de grandes dimensiones es costosa y el análisis de los resultados es muy complejo. Por el contrario, la *experimentación secuencial* consiste en emplear la información obtenida en un experimento (pequeño) para planificar el siguiente. Por ejemplo, la metodología a seguir puede ser la siguiente:

1. Explorar la respuesta, mediante *experimentos de screening*. Este tipo de experimentos es adecuado cuando existe un número muy elevado de factores, y se pretende en primer lugar identificar cuáles de estos factores tienen una influencia significativa sobre la respuesta (factores de “primer orden” o “clave”) y cuáles tienen una influencia despreciable. El objetivo es reducir el número de factores que deben ser considerados en posteriores experimentos.
2. Optimizar la respuesta, usando un *diseño experimental de segundo orden* (o diseño de la superficie de respuesta), con un reducido número de factores clave.
3. Realizar *experimentos de sensibilidad*, en los que se pretende explorar el rango de operación, por ejemplo variando el valor de los factores $\pm 10\%$ alrededor del punto óptimo.
4. Realizar *experimentos de confirmación*, con el fin de validar las conclusiones obtenidas.

Una forma de simplificar el planteamiento de un problema complejo consiste en descomponerlo en problemas de menor complejidad:

- Se pueden dividir los sistemas complejos en subsistemas, y realizar pequeños experimentos internos a los subsistemas. Para estudiar la interacción entre los subsistemas, se pueden plantear experimentos entre ellos.
- Se pueden dividir los procesos secuenciales en pasos y realizar pequeños experimentos sobre los procesos individuales de cada paso. Se pueden realizar experimentos entre los pasos, para analizar su interacción.

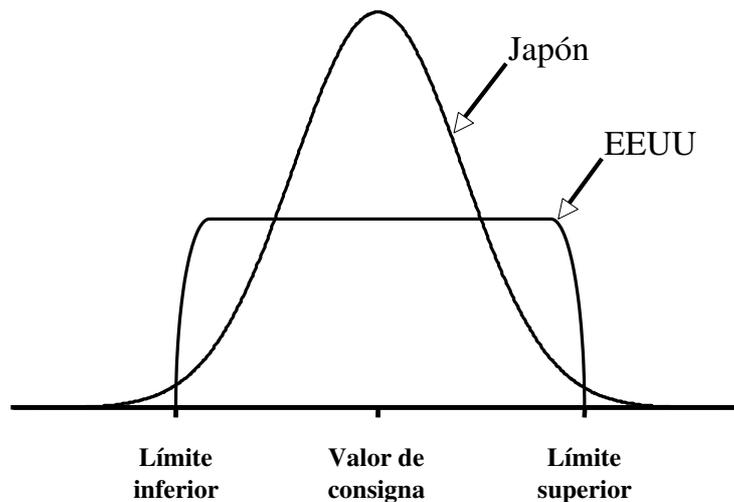


Figura 13.2: Distribución del tamaño de las piezas.

Selección de las respuestas

Las respuestas pueden tomar

- valores *numéricos continuos* (por ejemplo, tiempo de espera medio en una cola),
- valores *numéricos discretos* (por ejemplo, número total de entidades procesadas), y también pueden tomar
- valores *categoricos* (por ejemplo, el sistema pasa o falla determinado criterio).

Los objetivos que pueden plantearse con relación a la respuesta son de distintos tipos. Por ejemplo:

- *Centrar el valor de la respuesta* en un determinado valor de consigna. Por ejemplo, se desea que la producción media diaria de una cadena de montaje sea una determinada. Una producción mayor acarrea gastos de almacenamiento del producto no demandado. Por el contrario, una producción menor reduce los posibles ingresos y genera insatisfacción en el cliente.
- *Minimizar el valor de la respuesta*. Por ejemplo, se puede investigar la política de gestión del inventario que minimice el coste total mensual.
- *Maximizar el valor de la respuesta*. Por ejemplo, se puede investigar qué factores aumentan la producción media diaria de una fábrica, o qué factores aumentan el número medio de productos que diariamente superan determinada prueba de calidad.

Frecuentemente las respuestas del experimento son estadísticos, tales como:

- Medidas de la tendencia central, como son la media o la mediana.
- Medidas de la variabilidad, o dispersión de los datos, como son el rango, la varianza, la desviación estándar, etc.

La selección de las respuestas del experimento es un tema de suma importancia. El siguiente ejemplo pretende ilustrar este hecho.

Ejemplo 13.2.1. *En un estudio cuyo objetivo sea la mejora de la calidad en el proceso de fabricación, considerar como respuesta la fracción de la producción que superan una determinada prueba de calidad puede no ser una elección afortunada. Existen muestras de ello.*

Una conocida marca de automóviles producía transmisiones en su fábrica de EEUU y en su fábrica de Japón. Pese a que en ambas fábricas las especificaciones eran las mismas, las transmisiones fabricadas en EEUU eran considerablemente más ruidosas y generaban más

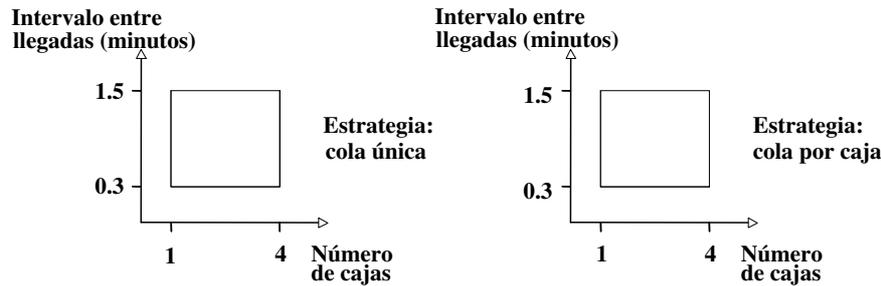


Figura 13.3: Región experimental.

reclamaciones durante su periodo de garantía que las fabricadas en Japón. La investigación del origen del problema reveló que:

- En la fábrica de EEUU, el proceso de fabricación se había optimizado atendiendo a una única respuesta: la fracción de piezas defectuosas. El objetivo planteado fue reducir la fracción de piezas defectuosas, y en efecto, todas las transmisiones fabricadas estaban dentro de especificaciones, pero no necesariamente próximas a su valor de consigna. Tal como se muestra en la Figura 13.2, el rango de la distribución de tamaños de las piezas estaba completamente contenido dentro de las especificaciones.
- Por el contrario, en la fábrica de Japón, se había optimizado el proceso de fabricación atendiendo a dos respuestas: el valor esperado y la variabilidad de la distribución de tamaños de las piezas. Como consecuencia, no todas las piezas estaban dentro de especificaciones, pero el tamaño de las piezas estaba más próximo al valor de consigna (ver Figura 13.2).

□

Selección de los factores experimentales

La selección de los factores se basa en el conocimiento sobre el sistema: se seleccionan como factores aquellos parámetros del modelo que previsiblemente influyen sobre la respuesta o permiten alcanzar los objetivos del experimento.

Generalmente es preferible incluir en las primeras fases del estudio demasiados factores que correr el riesgo de no considerar un factor importante.

Los valores de cada factor que van a ser considerados en el estudio (denominados *niveles*), deben escogerse de modo que estén lo suficientemente alejados como para producir un cambio significativo en la respuesta, pero tan alejados que exista un salto en la información de la respuesta entre los niveles. Normalmente es suficiente con escoger 2 niveles para cada factor. Atendiendo a sus valores, los factores pueden clasificarse en los dos tipos siguientes:

- *Cuantitativos*. Toman valores numéricos. Un factor cuantitativo sería el número de empleados en un banco, el número cajas de cobro en un supermercado, etc.
- *Cualitativos*. Normalmente cada posible valor de un factor cuantitativo representa una determinada forma de organización o disposición estructural del modelo. Por ejemplo, un factor cualitativo sería la disciplina de determinada cola: FIFO, LIFO, etc.

Ejemplo 13.2.2. En un sistema consistente en una cadena de montaje, posibles factores cuantitativos son el número de máquinas y de operarios, la velocidad de la cinta transportadora de la cadena, etc. Posibles factores cualitativos son la disciplina de las colas, la estrategia seguida al distribuir a los operarios en las máquinas, etc. □

Ejemplo 13.2.3. Se pretende evaluar el impacto, en el coste y en el tiempo medio de espera en cola, de dos cambios de funcionamiento de una línea de cajas en un supermercado. Los cambios consisten en variar el número de cajas y la reorganizar de las colas. La frecuencia de llegada de los clientes afecta significativamente al funcionamiento del sistema, por ello se incluye también como un factor experimental. Para cada uno de los tres factores experimentales anteriormente indicados, se han seleccionado dos niveles. En la Figura 13.3 se muestra la región experimental del estudio:

- Factor: tiempo medio entre llegadas sucesivas de los clientes. Niveles: 0.3 minutos (hora punta) y 1.5 minutos (hora valle).
- Factor: número de cajas. Niveles: 1 caja (valor mínimo posible) y 4 cajas (número máximo de cajas).
- Factor: estrategia de las colas. Niveles: establecer una cola única para todas las cajas o establecer una cola por caja.

□

Puede establecerse otra clasificación de los factores, complementaria a la anterior, en función de que se pueda o no ejercer una acción sobre ellos en “el mundo real”. Los factores pueden clasificarse en:

- *Manipulables*. Representan opciones modificables en el sistema real, por ejemplo, el número de empleados en un banco o la disciplina de una cola.
- *No manipulables*. Son aquellos factores sobre cuyo valor no se puede influir en el sistema real. Por ejemplo, la frecuencia de llegada de los clientes al banco o el volumen medio de demandas de un producto.

El que un factor sea manipulable o no depende de las circunstancias particulares del sistema bajo estudio. Normalmente, cuando se busca solución a un problema real, los factores experimentales suelen ser manipulables: interesa centrarse en el estudio de aquellos factores que pueden ser modificados en el sistema real.

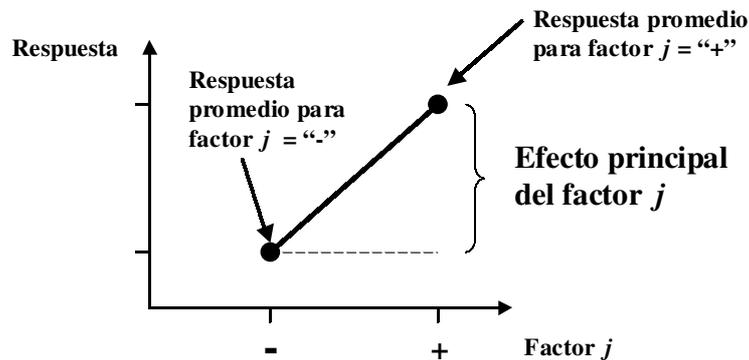
No obstante, los factores no manipulables también pueden ser de interés en un estudio, por ejemplo, puede interesar estudiar cómo un incremento abrupto en la frecuencia de llegada de entidades afecta a la congestión de un sistema de colas. No en vano, una de las grandes ventajas del modelado matemático y la simulación es que en un experimento de simulación se puede manipular a voluntad el valor de todas las variables, con independencia de que sean manipulables o no en el mundo real.

13.3. Diseño experimental factorial completo

EL objetivo es estudiar cómo k factores afectan a determinada medida de la respuesta del modelo y si existe *interacción* entre los factores (es decir, si el efecto de un factor depende del nivel de los otros). Para ello puede diseñarse un experimento *factorial completo*, también llamado *diseño factorial* 2^k . Esta técnica de diseño consta de los pasos siguientes:

1. Se seleccionan dos niveles para cada factor. Se representan “-” y “+”. En el caso de factores cuantitativos, se asigna “-” al valor numérico menor y “+” al mayor. La selección los niveles debe realizarla el ingeniero a partir de su conocimiento del sistema y de los objetivos del estudio.
2. Se construye la matriz del diseño, que contiene las combinaciones de niveles de los factores que deben simularse. Cada una de estas combinaciones recibe el nombre de *puntos del diseño*. La matriz de un diseño factorial completo consta de las 2^k posibles combinaciones de valores de los niveles (2 por factor).

Ejemplo 13.3.1. En la tabla adjunta se muestra la matriz del diseño para el caso $k = 3$. La variable R_i (con $i = 1, 2, \dots, 8$) representa el valor de la respuesta obtenido al simular el modelo con los niveles correspondientes al punto del diseño i . Por ejemplo, R_3 es la respuesta obtenida al simular el modelo con los factores 1 y 3 al nivel “-” y con el factor 2 al nivel “+”.

Figura 13.4: Efecto principal del factor j .

Como se verá más adelante, la escritura de la matriz de diseño facilita el cálculo de los efectos principales de los factores y de sus interacciones.

| Punto del diseño | Factor 1 | Factor 2 | Factor 3 | Respuesta R_i |
|------------------|----------|----------|----------|-----------------|
| 1 | - | - | - | R_1 |
| 2 | + | - | - | R_2 |
| 3 | - | + | - | R_3 |
| 4 | + | + | - | R_4 |
| 5 | - | - | + | R_5 |
| 6 | + | - | + | R_6 |
| 7 | - | + | + | R_7 |
| 8 | + | + | + | R_8 |

□

Efectos principales

El *efecto principal* del factor j se representa e_j , y es la diferencia entre la respuesta media cuando el factor j vale “+” y la respuesta media cuando el factor j vale “-”. En la Figura 13.4 se representa gráficamente el concepto de efecto principal del factor j . Los promedios se calculan sobre todas las posibles combinaciones de los restantes $k-1$ factores. Por tanto, para calcular e_j se aplican los signos de la columna “Factor j ” a las correspondientes respuestas, se realiza la suma y se divide por 2^{k-1} . A continuación se muestra un ejemplo.

Ejemplo 13.3.2. Para el diseño 2^3 factorial del Ejemplo 13.3.1:

$$\begin{aligned} \text{Efecto principal del factor 1: } e_1 &= \frac{-R_1 + R_2 - R_3 + R_4 - R_5 + R_6 - R_7 + R_8}{4} \\ \text{Efecto principal del factor 2: } e_2 &= \frac{-R_1 - R_2 + R_3 + R_4 - R_5 - R_6 + R_7 + R_8}{4} \\ \text{Efecto principal del factor 3: } e_3 &= \frac{-R_1 - R_2 - R_3 - R_4 + R_5 + R_6 + R_7 + R_8}{4} \end{aligned}$$

□

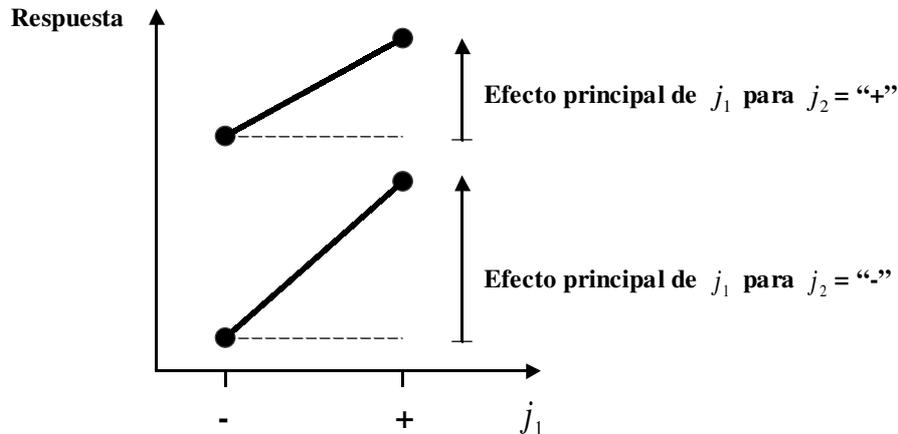


Figura 13.5: Gráfico de interacción de dos factores. Líneas no paralelas: existe interacción.

Interacciones de dos factores

Frecuentemente, el efecto sobre la respuesta de un factor, j_1 , depende del nivel de otro factor, j_2 . Se dice entonces que *existe una interacción* entre los factores j_1 y j_2 . Cuando la interacción entre dos factores es grande, los efectos principales de esos factores pierden su significado: no pueden ser interpretados sin considerar la interacción.

En los diseños experimentales con dos niveles por factor, la *interacción* entre dos factores, j_1 y j_2 , se representa e_{j_1, j_2} , y se define de la forma siguiente:

$$e_{j_1, j_2} = \frac{\left(\begin{array}{c} \text{Efecto principal de } j_1 \\ \text{cuando } j_2 = "+" \end{array} \right) - \left(\begin{array}{c} \text{Efecto principal de } j_1 \\ \text{cuando } j_2 = "-" \end{array} \right)}{2} \quad (13.1)$$

Las interacciones (también llamadas *efectos cruzados*) son simétricas:

$$e_{j_1, j_2} = e_{j_2, j_1} \quad (13.2)$$

Si se codifican los niveles "+" y "-" como +1 y -1 respectivamente, entonces e_{j_1, j_2} es la diferencia entre las pendientes de las dos líneas en el gráfico de interacción. En la Figura 13.5 se representa gráficamente el fenómeno de la interacción entre factores: los dos factores interactúan si las dos líneas en el gráfico de interacción no son paralelas.

Ejemplo 13.3.3. En un diseño 2^3 factorial, los efectos cruzados tienen la interpretación siguiente:

$$e_{1,2} = \frac{\left[\frac{R_4+R_8}{2} - \frac{R_3+R_7}{2} \right]}{1 - (-1)} - \frac{\left[\frac{R_2+R_6}{2} - \frac{R_1+R_5}{2} \right]}{1 - (-1)} \quad (13.3)$$

$$e_{1,3} = \frac{\left[\frac{R_6+R_8}{2} - \frac{R_5+R_7}{2} \right]}{1 - (-1)} - \frac{\left[\frac{R_2+R_4}{2} - \frac{R_1+R_3}{2} \right]}{1 - (-1)} \quad (13.4)$$

$$e_{2,3} = \frac{\left[\frac{R_7+R_8}{2} - \frac{R_5+R_6}{2} \right]}{1 - (-1)} - \frac{\left[\frac{R_3+R_4}{2} - \frac{R_1+R_2}{2} \right]}{1 - (-1)} \quad (13.5)$$

□

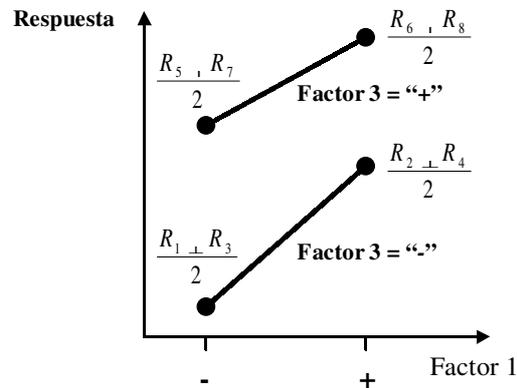


Figura 13.6: Interacción entre los factores 1 y 3.

Existe un procedimiento sencillo para escribir las fórmulas de los efectos cruzados. Consiste en crear nuevas columnas, $j_1 \times j_2$, en la matriz experimental, que contenga los signos obtenidos de multiplicar el signo de la columna j_1 por el correspondiente signo de la columna j_2 . Esta nueva columna contiene los signos que afectan a las respuestas R_i en la expresión e_{j_1, j_2} .

Ejemplo 13.3.4. La expresión de la interacción entre los factores 1 y 3 en un diseño 2^3 factorial puede obtenerse creando una nueva columna “1 × 3” en la matriz experimental, que contenga los signos obtenidos de multiplicar el signo de la columna 1 por el correspondiente signo de la columna 3. Esta nueva columna contiene los signos que afectan a las respuestas R_i en la expresión de $e_{1,3}$. Sumando las respuestas, cada una afectada por su correspondiente signo, y dividiendo por 2^{k-1} , se obtiene la expresión para el efecto cruzado. Análogamente se añaden las columnas “1 × 2” y “2 × 3”:

| Punto | Factor 1 | Factor 2 | Factor 3 | 1 × 2 | 1 × 3 | 2 × 3 | R_i |
|-------|----------|----------|----------|-------|-------|-------|-------|
| 1 | - | - | - | + | + | + | R_1 |
| 2 | + | - | - | - | - | + | R_2 |
| 3 | - | + | - | - | + | - | R_3 |
| 4 | + | + | - | + | - | - | R_4 |
| 5 | - | - | + | + | - | - | R_5 |
| 6 | + | - | + | - | + | - | R_6 |
| 7 | - | + | + | - | - | + | R_7 |
| 8 | + | + | + | + | + | + | R_8 |

A partir de la tabla anterior, se obtienen las siguientes expresiones:

$$e_{1,2} = \frac{R_1 - R_2 - R_3 + R_4 + R_5 - R_6 - R_7 + R_8}{4} \tag{13.6}$$

$$e_{1,3} = \frac{R_1 - R_2 + R_3 - R_4 - R_5 + R_6 - R_7 + R_8}{4} \tag{13.7}$$

$$e_{2,3} = \frac{R_1 + R_2 - R_3 - R_4 - R_5 - R_6 + R_7 + R_8}{4} \tag{13.8}$$

□

Los efectos principales pueden obtenerse a partir de la gráfica de interacción. En la Figura 13.7 se muestra el procedimiento: el gráfico de los efectos principales es la media de los correspondientes niveles del gráfico de interacción.

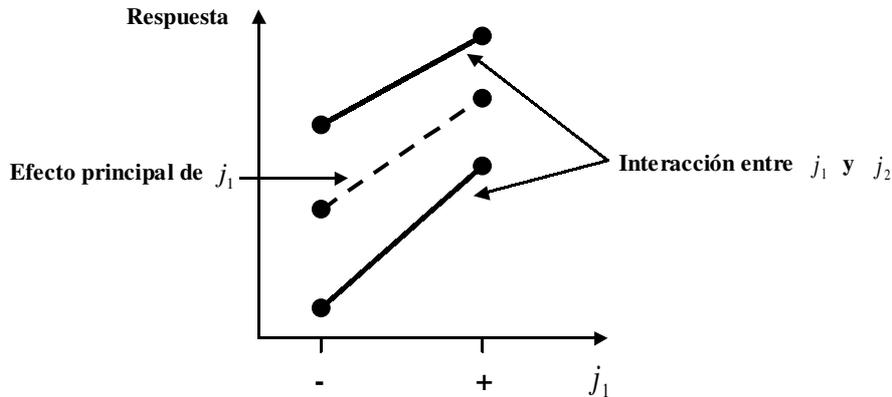


Figura 13.7: Obtención de los efectos principales a partir del gráfico de interacción.

Ejemplo 13.3.5. *Considérese el modelo de la gestión del inventario de un único producto. Los dos parámetros que condicionan la gestión del inventario son s y S :*

- *El parámetro s determina el nivel del inventario por debajo del cual se realiza el pedido.*
- *El parámetro S determina el nivel de inventario que se pretende alcanzar realizando el pedido.*

Para realizar el estudio es conveniente definir un tercer parámetro, d , de la forma siguiente:

$$d = S - s \tag{13.9}$$

El objetivo del estudio es analizar cómo afectan los factores s y d al coste total mensual. Se escogen dos niveles por factor, con los valores siguientes:

| | | |
|-----|----|----|
| | - | + |
| s | 20 | 60 |
| d | 10 | 50 |

Una vez realizadas las simulaciones de los cuatro puntos experimentales, los resultados obtenidos son los siguientes:

| Punto experimental | s | d | $s \times d$ | Respuesta |
|--------------------|-----|-----|--------------|-----------|
| 1 | - | - | + | 141.86 |
| 2 | + | - | - | 141.37 |
| 3 | - | + | - | 112.45 |
| 4 | + | + | + | 146.52 |

Los efectos principales y la interacción de los dos factores son (ver Figura 13.8):

$$e_s = \frac{-141.86 + 141.37 - 112.45 + 146.52}{2} = 16.79 \tag{13.10}$$

$$e_d = \frac{-141.86 - 141.37 + 112.45 + 146.52}{2} = -12.13 \tag{13.11}$$

$$e_{s,d} = \frac{146.52 - 112.45}{2} - \frac{141.37 - 141.86}{2} = 17.05 - (-0.24) = 17.28 \tag{13.12}$$

Así pues, pueden extraerse las siguientes conclusiones:

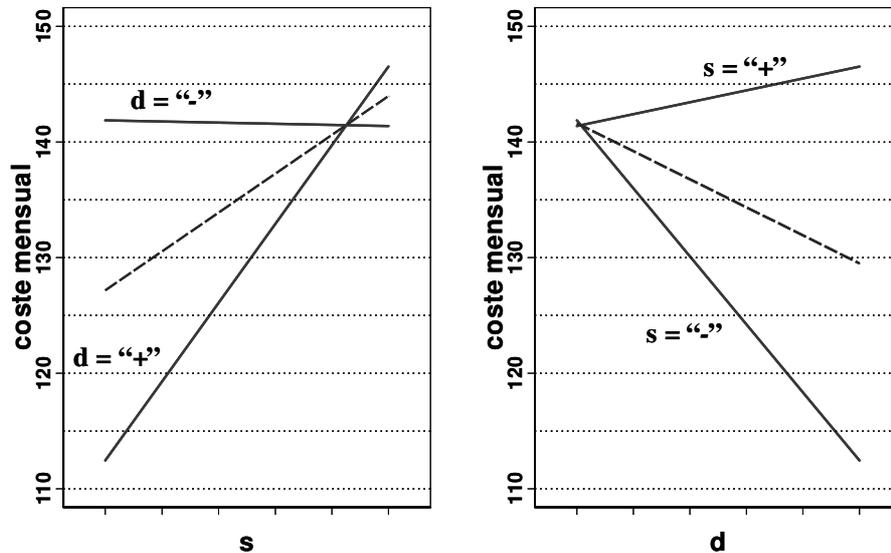


Figura 13.8: Efectos principales (líneas de puntos) e interacción.

- El efecto promedio de aumentar s de 20 a 60 es aumentar el coste medio mensual en 16.79.
- El efecto promedio de aumentar d de 10 a 50 es disminuir el coste medio mensual en 12.13.
- Sin embargo, la interacción entre los dos factores es tan fuerte que las conclusiones anteriores, extraídas de los efectos principales, no son correctas. Obsérvese, por ejemplo, que en función del valor de s , un aumento en d puede producir un incremento o una reducción del coste. Igualmente, cuando d está a su nivel inferior, el coste es prácticamente insensible al nivel de s . Por el contrario, cuando d está a su nivel superior, el coste aumenta drásticamente al aumentar s .

□

Interacción entre más de dos factores

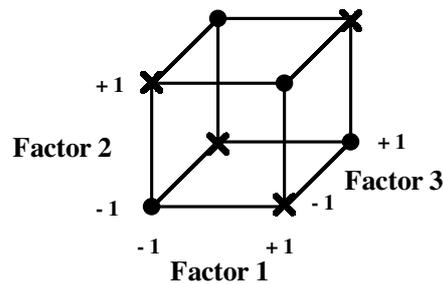
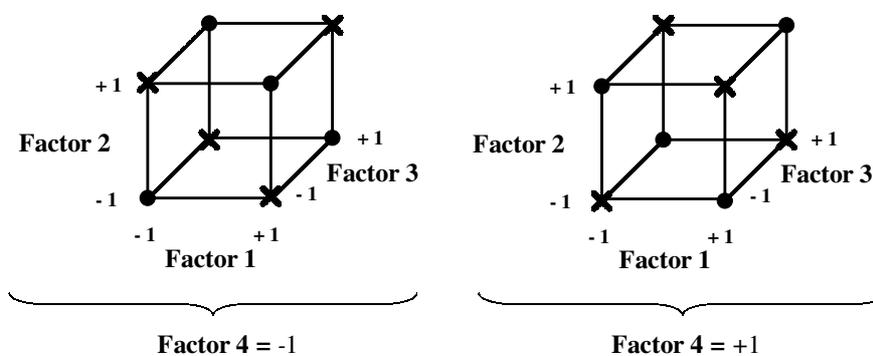
Tres factores interactúan si la interacción entre dos factores depende del nivel del tercer factor. Normalmente las interacciones entre tres o más factores no son importantes, y se suele suponer que son despreciables. Estas interacciones de tercer orden y superiores corresponden con términos de tercer orden y superiores en el desarrollo de Taylor de la respuesta en función de los factores.

Intervalos de confianza

Puesto que las respuestas R_i son variables aleatorias, los efectos también son aleatorios. Para contrastar si las diferencias observadas entre los efectos son "reales" o, por el contrario, son únicamente debidas a las fluctuaciones estadísticas, es preciso estimar las varianzas de los efectos.

Para ello, puede replicarse independientemente la simulación de cada punto del diseño n veces, de modo que se obtengan n observaciones independientes de cada efecto. A continuación se muestra un ejemplo.

Ejemplo 13.3.6. Se repite $n = 10$ veces el experimento 2^2 factorial del sistema de gestión del inventario del Ejemplo 13.3.5. Con ello, se obtienen 10 observaciones independientes de cada

Figura 13.9: Diseño factorial fraccional 2^{3-1} .Figura 13.10: Diseño factorial fraccional 2^{4-1} .

uno de los tres efectos: e_s , e_d y $e_{s,d}$. De estas muestras de 10 observaciones, pueden construirse intervalos del 90% de confianza para la media de los factores. Los valores obtenidos son:

- Intervalo con un 90% de confianza para $E\{e_s\}$: 17.66 ± 1.02 .
- Intervalo con un 90% de confianza para $E\{e_d\}$: -8.77 ± 1.27 .
- Intervalo con un 90% de confianza para $E\{e_{s,d}\}$: 10.60 ± 1.94 .

A la vista de los intervalos para la media de los efectos, parece que todos los efectos son “reales”. Otro tema diferente es decidir si los valores numéricos obtenidos para los efectos son considerables desde el punto de vista práctico. Es decir, si los efectos, aunque estadísticamente significativos, son muy pequeños, en la práctica se considerará que el efecto de realizar los cambios estudiados en la política de gestión del inventario es despreciable. \square

13.4. Diseños experimentales factoriales fraccionales

En las primeras fases de la experimentación, el número de factores que se desea investigar suele ser grande, con lo cual, el número de puntos experimentales del diseño factorial completo es muy elevado. En ocasiones es tan elevado que es irrealizable. Este suele ser el caso de los experimentos denominados de “screening”, en los que están presentes muchos factores y se desea determinar, con un esfuerzo razonable, qué factores influyen de manera relevante sobre la respuesta y cuáles tienen un efecto tan pequeño que puede considerarse despreciable.

Una forma de proceder en estos casos consiste en eliminar de la matriz experimental un determinado conjunto de puntos experimentales. Este tipo de diseños experimentales se denominan *diseños factoriales fraccionales*.

Una familia importante de los *experimentos factoriales fraccionales* son los diseños 2^{k-p} , en los que se escogen (de manera balanceada) 2^{k-p} puntos de entre los 2^k de que consta el diseño factorial completo.

Ejemplo 13.4.1. En la Figura 13.9 se muestra el diseño experimental de un diseño factorial fraccional 2^{3-1} . Consta de 4 puntos de diseño, en lugar de 8: puede escogerse equivalentemente o los puntos señalados por un círculo o los señalados por un aspa. □

Ejemplo 13.4.2. Equivalentemente, se diseña el experimento factorial fraccional 2^{4-1} . En la Figura 13.10 se muestra el diseño experimental: pueden escogerse los puntos señalados con un círculo o los puntos señalados por un aspa. □

El precio que debe pagarse al realizar un experimento factorial fraccional es que disminuye el número de interacciones entre factores que pueden estimarse. No obstante, el experimento fraccional se diseña de modo que sea posible estimar los términos de menor orden (típicamente los efectos principales y las interacciones entre dos factores).

Ejemplo 13.4.3. Se realiza una modificación en el modelo de gestión del inventario, consistente en introducir los dos nuevos factores siguientes:

- El número de meses, m , que transcurren entre las sucesivas evaluaciones del inventario. En el modelo original $m = 1$, pero desea estudiarse el efecto de fijar $m = 3$.
 - Un nuevo factor surge debido a que el suministrador ha introducido una opción “*exprés*”.
 - Originalmente, si se ordenan Z unidades de producto, el coste es $3200 + 300 \cdot Z$ y el tiempo que tardará en recibirse el producto es una variable aleatoria distribuida $U(0.5, 1.0)$.
 - Si se utiliza el servicio “*exprés*”, el producto se recibirá en la mitad de tiempo: el tiempo que transcurre desde que se ordena el pedido hasta que se recibe es una variable aleatoria $U(0.25, 0.5)$; a cambio, el coste del servicio “*exprés*” es $4800 + 400 \cdot Z$.
- Así pues, se introduce el factor cualitativo “*prioridad del pedido*”, P , con dos niveles: “*normal*” o “*exprés*”.

La tabla de codificación de este modelo, de $k = 4$ parámetros, es la siguiente:

| | | |
|-----|--------|--------|
| | – | + |
| s | 20 | 60 |
| S | 70 | 120 |
| m | 1 | 3 |
| P | Normal | Exprés |

Un diseño completo 2^4 factorial tiene 16 puntos experimentales. Sin embargo, si el objetivo es únicamente estimar los efectos principales, puede realizarse un experimento factorial fraccional 2^{4-1} (con este experimento no pueden estimarse las interacciones entre las parejas de factores).

Se realiza una simulación de cada uno de los 8 puntos experimentales, obteniéndose los resultados siguientes:

| Punto experimental | s | S | m | P | Respuesta |
|--------------------|-----|-----|-----|-----|-----------|
| 1 | – | – | – | – | 118.686 |
| 2 | + | – | – | + | 196.119 |
| 3 | – | + | – | + | 166.674 |
| 4 | + | + | – | – | 149.542 |
| 5 | – | – | + | + | 148.045 |
| 6 | + | – | + | – | 135.754 |
| 7 | – | + | + | – | 153.202 |
| 8 | + | + | + | + | 181.885 |

Los efectos se calculan del mismo modo que en los diseños 2^k factoriales; por ejemplo, el efecto principal de s se calcula aplicando los signos de la columna "s" a las correspondientes respuestas, realizando la suma y dividiendo el resultado por $2^{k-p-1} = 4$. Se obtiene:

$$e_s = 19.173 \quad e_S = 13.175 \quad e_m = -3.034 \quad e_p = 33.885 \quad (13.13)$$

Para estimar las varianzas de los efectos principales, se replica el experimento $n = 10$ veces y se construye un intervalo del 90% de confianza para el valor esperado de cada efecto principal:

$$\begin{array}{ll} \text{Para E}(e_s): & 15.478 \pm 1.831 \\ \text{Para E}(e_S): & 14.431 \pm 2.017 \\ \text{Para E}(e_m): & -1.217 \pm 1.449 \\ \text{Para E}(e_p): & 34.915 \pm 1.775 \end{array}$$

Las conclusiones son las siguientes:

- Los efectos principales de s y S son significativamente positivos.
- El efecto principal de m es muy pequeño y estadísticamente insignificante, lo cual indica que no hay que incluirlo como factor en futuras simulaciones, al menos, en el rango en que hemos especificado sus niveles.
- La opción "expres" aumenta considerablemente el coste de operación, con lo cual parece que no compensa.

□

13.5. Optimización

En los estudios de simulación es frecuente realizar en primer lugar experimentos factoriales completos o fraccionales, con el fin de identificar cuáles son los factores de primer orden. A continuación, se busca qué niveles de estos factores clave optimizan la respuesta del sistema. Cuando los factores son cuantitativos, puede emplearse la *metodología de la superficie de respuesta* para resolver este problema de optimización. La *superficie de respuesta* es el valor esperado de la respuesta en función de los factores. A continuación de muestra un ejemplo de aplicación de esta metodología.

Ejemplo 13.5.1. *Considérese el modelo original de la gestión del inventario, con 2 factores: s y S . El estudio del sistema comienza con un diseño factorial 2^2 . Las 4 combinaciones de los factores (s, S) :*

$$(-, -) \quad (+, -) \quad (-, +) \quad (+, +)$$

constituyen los vértices de un cuadrado en el plano (s, S) .

Se realizan 10 réplicas independientes de la simulación de cada punto experimental, obteniéndose los siguientes valores medios de la respuesta:

$$\begin{array}{ll} (-, +): R = 133.114 & (+, +): R = 151.682 \\ (-, -): R = 119.684 & (+, -): R = 144.194 \end{array}$$

Como una aproximación lineal de la superficie de respuesta cubierta por este diseño (el área comprendida dentro del cuadrado), se puede ajustar por mínimos cuadrados el modelo lineal

$$R = \beta_0 + \beta_1 \cdot s + \beta_2 \cdot S \quad (13.14)$$

a la respuesta en los 4 puntos de diseño. Realizando los cálculos, se obtienen los siguientes coeficientes del ajuste por mínimos cuadrados:

$$\hat{\beta}_0 = 95.757 \quad \hat{\beta}_1 = 0.538 \quad \hat{\beta}_2 = 0.209 \quad (13.15)$$

Dado que en este caso la respuesta es el coste asociado a la gestión del inventario, interesa minimizar la respuesta. Para ello, se examina el modelo lineal ajustado para determinar qué dirección en el espacio (s, S) es la de mayor decrecimiento de la respuesta.

Dada una función $f(x_1, \dots, x_n)$, el valor negativo del vector de las derivadas parciales de la función

$$-\left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) \quad (13.16)$$

particularizadas en un punto, apunta en la dirección de más rápido descenso de la función en ese punto.

En el caso del modelo lineal representado por la Ecuación (13.14), la dirección de más rápido descenso de la respuesta en el plano (s, S) , es paralela al vector

$$-\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial S}\right) \quad (13.17)$$

es decir, a

$$\left(-\hat{\beta}_1, -\hat{\beta}_2\right) \quad (13.18)$$

Así pues, la búsqueda del valor mínimo de la respuesta debe desplazarse, desde el punto central del cuadrado, por el plano (s, S) , en la dirección $(-0.538, -0.209)$.

La búsqueda se desplaza realizando simulaciones para (s, S) , avanzando en la dirección del vector, hasta que se observe que la respuesta comience a crecer. Entonces, se retrocede hasta la respuesta anterior menor y se usa como centro para un nuevo diseño 2^2 factorial, del cual obtenemos un nuevo ajuste lineal y una dirección de más rápido descenso.

Cuando el ajuste lineal sea aproximadamente plano, es decir, cuando $\hat{\beta}_1$ y $\hat{\beta}_2$ estén ambos próximos a cero, es indicación de que se está cerca de un mínimo (posiblemente, sólo local). En este punto, puede realizarse un diseño experimental más elaborado que el factorial, que permita realizar el ajuste a un modelo cuadrático:

$$R = \beta_0 + \beta_1 \cdot s + \beta_2 \cdot S + \beta_{12} \cdot s \cdot S + \beta_{11} \cdot s^2 + \beta_{22} \cdot S^2 \quad (13.19)$$

del cual se calculará el punto mínimo. □

13.6. Optimización usando Arena

Arena trae incorporada una herramienta de optimización denominada OptQuest, cuya finalidad es explorar automáticamente qué valores de las entradas de control producen un valor extremo (máximo o mínimo, según se indique) de la respuesta. OptQuest permite definir restricciones sobre las entradas de control. Es decir, delimitar el conjunto de valores sobre los que se realiza la búsqueda.

Para realizar un estudio de optimización usando OptQuest, primeramente hay que arrancar Arena y abrir ("File / Open") el modelo que se desee estudiar. A continuación, hay que arrancar OptQuest, desde la ventana de Arena. Para ello hay que seleccionar: "Tools / Optquest for Arena". OptQuest trabaja sobre el modelo previamente abierto en Arena.

Una vez abierta la ventana de OptQuest, hay que iniciar una nueva sesión: "File / New". Con ello aparece una ventana en la que se muestran los potenciales controles (o factores, como se prefiera) que hay en el modelo. Estos son: la capacidad de los recursos y las variables.

| Control Selection: Select controls and set bounds. | | | | | | | |
|--|--------------------------|--------------------------|-------------|-----------------|-------------|--------------|----------|
| | Select | Control | Lower Bound | Suggested Value | Upper Bound | Type | Category |
| | <input type="checkbox"/> | asiento | 29 | 30 | 31 | Discrete (1) | Resource |
| | <input type="checkbox"/> | empleado peticion y pago | 5 | 6 | 7 | Discrete (1) | Resource |
| | <input type="checkbox"/> | empleado recogida | 1 | 2 | 3 | Discrete (1) | Resource |

Figura 13.11: Posibles controles del modelo.

| Control Selection | | | | | | | |
|-------------------|-------------------------------------|--------------------------|-------------|-----------------|-------------|--------------|----------|
| | Select | Control | Lower Bound | Suggested Value | Upper Bound | Type | Category |
| | <input checked="" type="checkbox"/> | asiento | 20 | 24 | 40 | Discrete (2) | Resource |
| | <input checked="" type="checkbox"/> | empleado peticion y pago | 1 | 4 | 7 | Discrete (1) | Resource |
| | <input checked="" type="checkbox"/> | empleado recogida | 1 | 4 | 7 | Discrete (1) | Resource |

Figura 13.12: Configuración de los controles del modelo.

Ejemplo 13.6.1. *Considérese de nuevo el modelo del restaurante de comida rápida descrito en el Problema 6.5. El modelo contiene tres recursos: los empleados encargados de recibir las peticiones y de cobrar, los empleados que entregan la comida, y los asientos del comedor. OptQuest muestra los tres recursos como posibles controles (ver la Figura 13.11).*

De entre los posibles controles, deben seleccionarse aquellos que se deseen estudiar, señalando las correspondientes casillas de la columna “Select”. Para estos controles seleccionados, debe especificarse (ver la Figura 13.11):

- El rango de valores que se desea estudiar, que es el comprendido entre “Lower Bound” y “Upper Bound”.
- El valor del control por el cual debe comenzar la optimización: “Suggested Value”. Cuanto más próximo esté este valor al valor solución, más rápida se produce la convergencia, y más garantías hay de que la búsqueda converge a la solución.
- El tipo del control (“Type”) puede ser continuo o discreto. En ocasiones los controles discretos debe especificarse también el tamaño de su paso de variación (“Input Step Size”). Este es el valor en el cual OptQuest va incrementando el control.

Ejemplo 13.6.2. *Los controles del Ejemplo 13.6.1 se han definido según se muestra en la Figura 13.12. En total hay 8 empleados, con lo cual el número de empleados que puede haber en cada mostrador varía entre 1 y 7. En el comedor hay 2 asientos por mesa, con lo cual el número de asientos debe variarse en pasos de 2. El número mínimo de asientos es 20 y el número máximo 40.* □

Una vez definidos los controles, hay que pulsar “OK” en la ventana “Control Selection”. Se abre una nueva ventana (“Constraints”), en la cual deben indicarse las restricciones que se imponen. Una vez definidas las restricciones, pulsar “OK”.

Ejemplo 13.6.3. *Prosiguiendo con el Ejemplo 13.6.2, la ligadura que debe imponerse es que la suma del número de empleados en los dos mostradores debe ser igual a 8. En la Figura 13.13 se muestra la restricción.* □

A continuación se abre la ventana “Objective and Requirement Selection”, en la que OptQuest muestra todas las posibles respuestas del modelo. Debe definirse el papel de cada una

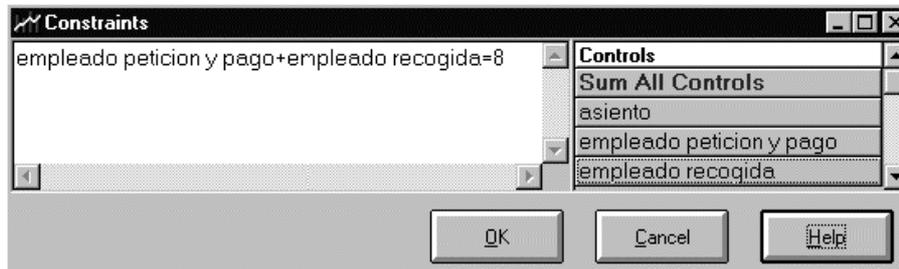


Figura 13.13: Restricción.

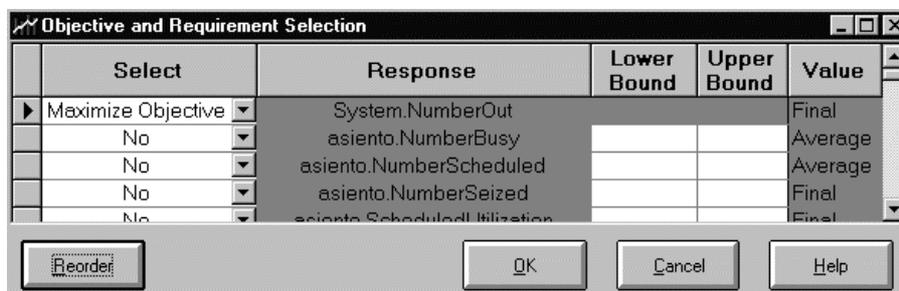


Figura 13.14: Respuesta.

de estas respuestas en el estudio de optimización. Para ello, debe seleccionarse una de las siguientes opciones:

- *No*: es la opción por defecto, e indica que esa respuesta no es relevante para el estudio de optimización.
- *Maximize Objective* o *Minimize Objective*: respuesta cuyo valor se pretende maximizar o minimizar respectivamente.
- *Requirement*: se especifica un requisito que debe satisfacer esta respuesta. Es análogo a las restricciones, pero en lugar de aplicado a los controles, aplicado a las respuestas.

Ejemplo 13.6.4. Siguiendo con el Ejemplo 13.6.3, supóngase que la respuesta que se desea maximizar es el número de clientes que abandonan el sistema (ver la Figura 13.14). El botón “Reorder” muestra las respuestas seleccionadas en la parte superior de la lista. □

Una vez definidas las respuestas, se abre la ventana “Options”, donde se establece la forma en que OptQuest debe realizar la búsqueda. La ventana contiene varias lengüetas.

- *“Time”*. Se especifica la duración de la búsqueda, bien fijando el número de réplicas de la simulación, o bien el tiempo que debe durar la búsqueda. Este último tiempo no se refiere al tiempo simulado, sino al tiempo de funcionamiento del ordenador.
- *“Precision”*. Indica el número de réplicas de la simulación que deben realizarse para cada configuración experimental. Puede realizarse sólo una, o bien realizarse varias y estimar intervalos de confianza.

Ejemplo 13.6.5. Completando el Ejemplo 13.6.4, se ha seleccionado que la optimización se ejecute durante 15 minutos, y que el número de réplicas de cada configuración experimental varíe entre 3 y 10. Seleccionando “Run / Start”, comienza la ejecución de la optimización. Una vez finalizada la ejecución, OptQuest muestra el resultado obtenido. En este caso, la configuración para la cual un mayor número de clientes abandonan el sistema es asignar 6

empleados al mostrador de petición y pago y 2 al mostrador de recogida. En número de asientos obtenido es 24, es decir, coincide con el valor asignado inicialmente.

EJERCICIOS

13.1 Se ha diseñado un experimento factorial completo con el fin de estudiar el efecto de los factores A, B y C sobre la respuesta S. Además, se han añadido dos réplicas del punto central de la región experimental, con el fin de estimar la curvatura de la respuesta. El motivo de realizar dos réplicas del punto central, en lugar de sólo una, es estimar la dispersión de la respuesta. La matriz experimental y los resultados del experimento son los siguientes:

| A | B | C | S |
|-----|----|-----|-------|
| 60 | 5 | 550 | 10.93 |
| 60 | 5 | 700 | 10.19 |
| 60 | 15 | 550 | 7.17 |
| 60 | 15 | 700 | 6.94 |
| 180 | 5 | 550 | 19.61 |
| 180 | 5 | 700 | 17.50 |
| 180 | 15 | 550 | 12.46 |
| 180 | 15 | 700 | 11.77 |
| 120 | 10 | 625 | 11.61 |
| 120 | 10 | 625 | 11.17 |

Dibuje los gráficos de los efectos principales y de las interacciones. Extraiga conclusiones.

13.2 Considere las 8 primeras observaciones experimentales mostradas en el Problema 13.1, es decir, los resultados del experimentos 2^3 , y no considere las dos réplicas correspondientes al punto central. A partir de las 8 observaciones, calcule los coeficientes de un modelo compuesto por un término constante, tres términos de efectos principales, dos términos de interacción de los factores dos a dos, y un término de interacción de los tres factores. Es decir:

$$\begin{aligned}
 y_S &= \beta_0 \\
 &+ \beta_A \cdot x_A + \beta_B \cdot x_B + \beta_C \cdot x_C \\
 &+ \beta_{AB} \cdot x_A \cdot x_B + \beta_{AC} \cdot x_A \cdot x_C + \beta_{BC} \cdot x_B \cdot x_C \\
 &+ \beta_{ABC} \cdot x_A \cdot x_B \cdot x_C
 \end{aligned}$$

Para ello realice la siguiente normalización. Asigne el valor +1 a x_i cuando el factor i se encuentre en el nivel superior, y el valor -1 cuando se encuentre en el nivel inferior.

¿Qué relación existe entre el término β_0 y el promedio de las observaciones?

¿Qué relación existe entre los coeficientes β_A , β_B , β_C , y los efectos principales?

¿Qué relación existe entre los coeficientes β_{AB} , β_{AC} , β_{BC} y los efectos de interacción?

13.3 ¿Qué tipo de modelo se obtiene mediante un experimento factorial completo 2^2 ? Como en el Problema 13.2, considere que cada factor tiene dos niveles: +1 y -1. Construya la matriz experimental y el modelo.

13.4 ¿Qué tipo de modelo se obtiene de un experimento factorial completo 2^4 ?

13.5 Construya la matriz experimental de cada uno de los siguientes experimentos:

- Experimento factorial completo, con dos factores: A y B. El factor A tiene cuatro niveles: 30, 50, 80 y 90. El factor B tiene tres niveles: 5, 10 y 30.
- Experimento 2^3 , con tres factores: A, B, y C. Cada factor tiene dos niveles. Los niveles de A son 2900 y 4100. Los niveles de B son 925 y 975. Los de C son: 4.0 y 4.8.
- Un experimento factorial completo 2^2 , con tres réplicas del punto central. Los niveles del factor A son 10 y 20. Los del factor B son 170 y 185.

13.6 Considere de nuevo el experimento factorial completo 2^3 del Problema 13.2. Suponga que no es posible realizar las 8 réplicas de la simulación, y que se opta por hacer un experimento factorial fraccional 2^{3-1} . Indique cómo diseñaría el experimento. Con este tipo de experimento, ¿pueden estimarse todos los términos del modelo mostrado en el Problema 13.2?

13.7 Se ha realizado un experimento factorial fraccional 2^{4-1} . La matriz experimental obtenida es la siguiente:

| A | B | C | D | S |
|-----|-----|----|----|--------|
| 335 | 7.6 | 36 | 72 | 97.54 |
| 335 | 7.6 | 44 | 88 | 101.46 |
| 335 | 9.4 | 36 | 88 | 92.18 |
| 335 | 9.4 | 44 | 72 | 82.25 |
| 415 | 7.6 | 36 | 88 | 83.31 |
| 415 | 7.6 | 44 | 72 | 79.13 |
| 415 | 9.4 | 36 | 72 | 68.09 |
| 415 | 9.4 | 44 | 88 | 78.20 |

Ajuste un modelo compuesto por la suma de un término constante, los cuatro efectos principales y los 6 términos de interacción de los factores dos a dos. Es decir:

$$\begin{aligned}
 y_S &= \beta_0 \\
 &+ \beta_A \cdot x_A + \beta_B \cdot x_B + \beta_C \cdot x_C + \beta_D \cdot x_D \\
 &+ \beta_{AB} \cdot x_A \cdot x_B + \beta_{AC} \cdot x_A \cdot x_C + \beta_{AD} \cdot x_A \cdot x_D \\
 &+ \beta_{BC} \cdot x_B \cdot x_C + \beta_{BD} \cdot x_B \cdot x_D + \beta_{CD} \cdot x_C \cdot x_D
 \end{aligned}$$

¿Pueden estimarse todos los términos del modelo?

13.8 Un estudio de optimización, con dos factores (A y B) y una respuesta (S), se realizó siguiendo la metodología de la superficie de respuesta, en particular un diseño central compuesto. En primer lugar se realizó un experimento factorial completo 2^2 , con un punto central:

| A | B |
|-----|----|
| 120 | 20 |
| 120 | 30 |
| 180 | 20 |
| 180 | 30 |
| 150 | 25 |

El punto central se añadió con el fin de analizar si la respuesta tenía curvatura. ¿Cómo puede analizarse gráficamente si la respuesta tiene curvatura?

En efecto, se comprobó que la curvatura de la respuesta era significativa. Por ello, para determinar el punto extremo (máximo o mínimo) de la respuesta, es necesario emplear un modelo con términos cuadráticos. Con este fin, se añadieron cuatro puntos experimentales adicionales, con lo que diseño completo quedó de la forma siguiente:

| punto | A | B |
|-------|-----|----|
| 1 | 120 | 20 |
| 2 | 120 | 30 |
| 3 | 180 | 20 |
| 4 | 180 | 30 |
| 5 | 100 | 25 |
| 6 | 200 | 25 |
| 7 | 100 | 17 |
| 8 | 200 | 33 |
| 9 | 150 | 25 |

Dibuje la región experimental del experimento. ¿Permite este diseño ajustar un modelo de segundo orden, que sea suma de un término constante, efectos principales (A, B),

interacción ($A \cdot B$) y términos cuadráticos (A^2 , B^2)? ¿Cómo puede estimarse el valor extremo de la respuesta usando este modelo?

- 13.9** Se ha realizado un experimento factorial completo 2^3 , con dos réplicas del punto central. Se han medido tres respuestas: S_1 , S_2 y S_3 . La matriz experimental es la siguiente:

| A | B | C | S_1 | S_2 | S_3 |
|-----|-----|-----|-------|-------|-------|
| 60 | 5 | 550 | 10.93 | 2710 | 11.7 |
| 60 | 5 | 700 | 10.19 | 3233 | 10.8 |
| 60 | 15 | 550 | 7.17 | 3021 | 9.0 |
| 60 | 15 | 700 | 6.94 | 3638 | 8.1 |
| 180 | 5 | 550 | 19.61 | 2903 | 13.0 |
| 180 | 5 | 700 | 17.50 | 3679 | 12.2 |
| 180 | 15 | 550 | 12.46 | 3029 | 10.3 |
| 180 | 15 | 700 | 11.77 | 3814 | 9.3 |
| 120 | 10 | 625 | 11.61 | 3378 | 10.3 |
| 120 | 10 | 625 | 11.17 | 3295 | 11.1 |

Analice gráficamente los resultados del experimento y extraiga conclusiones.

- 13.10** Se pretende estudiar el funcionamiento de un proceso de montaje de ventiladores de techo. Las piezas necesarias para montar el ventilador llegan en kits. El intervalo de tiempo entre la llegada de los kits está distribuido TRIA (2, 5, 10) (todos los tiempos están expresados en minutos). Hay 4 operarios encargados de realizar el ensamblaje de las piezas. Cuando se recibe un kit, éste es enviado automáticamente al primer operario de la línea de ensamblaje que se encuentra libre. El tiempo que tarda cada operario en montar el ventilador sigue la distribución siguiente:

| Operario | Tiempo de ensamblaje |
|----------|----------------------|
| 1 | TRIA (15, 18, 20) |
| 2 | TRIA (16, 19, 22) |
| 3 | TRIA (16, 20, 24) |
| 4 | TRIA (17, 20, 23) |

Una vez montados, los ventiladores son revisados, encontrándose que el 7% de ellos están defectuosos. Un ventilador que está defectuoso es enviado de nuevo al operario que lo ha montado para que lo repare. Los ventiladores defectuosos tienen prioridad frente al montaje de nuevos ventiladores. Dado que para reparar el ventilador hay que desmontarlo y volverlo a montar, se supone que el tiempo de reparación es un 30% mayor que el tiempo normal de ensamblaje. Simule el funcionamiento del sistema durante 20000 minutos y estime la utilización de cada operario y el tiempo de ciclo de las entidades.

Suponga que puede contratar a un operario más, y que ese operario puede tener prestaciones idénticas a uno de los cuatro operarios ya existentes. ¿A cuál de ellos convendría escoger? Use PAN con 5 réplicas por escenario, y seleccione el escenario con un tiempo de ciclo menor. ¿Resulta sorprendente el resultado obtenido?

Suponga que es posible contratar hasta 5 nuevos operarios, y que cada uno de ellos puede tener prestaciones iguales a cualquiera de los 4 operarios existentes. Puede realizar la asignación como desee, incluyendo asignar los 5 nuevos empleados a un mismo operario. ¿Cuál es la mejor opción? Use OptQuest para buscar la colocación óptima para estos nuevos empleados (5 como máximo), con el objetivo de reducir el tiempo de ciclo.

Parte V

Apéndices

Apéndice A

Algunos conceptos básicos de probabilidad y estadística

Objetivos: Una vez estudiado el contenido del apéndice, debería:

- Haber “refrescado” los conceptos básicos de probabilidad y estadística necesarios para llevar a cabo estudios de simulación.

A.1. ¿Para qué?

La aplicación de conocimientos de probabilidad y estadística es necesaria en distintas fases de un proyecto de simulación. Algunas de estas fases son:

- La selección de las distribuciones de las entradas de la simulación.
- La generación de variables aleatorias de acuerdo con estas distribuciones.
- La validación del modelo.
- El análisis estadístico de los resultados de la simulación.
- El diseño de los experimentos a simular.

En este apéndice se introducen algunos conceptos básicos de probabilidad y estadística que son relevantes en todo estudio de simulación.

A.2. El significado de “probabilidad”

La probabilidad de un evento A es un número, $\text{Prob}\{A\}$, que se asigna al evento y que puede interpretarse de la manera siguiente:

Si el experimento se realiza n veces y el evento A ocurre n_A veces, entonces puede afirmarse *con un alto grado de certeza* que la frecuencia relativa $\frac{n_A}{n}$ con que sucede el evento A está *próxima* a $\text{Prob}\{A\}$ siempre que n sea *suficientemente grande*.

Esta interpretación es imprecisa: los términos “con alto grado de certeza”, “próxima” y “suficientemente grande” no tienen un significado claro. Sin embargo, no puede evitarse

esta falta de precisión. A continuación se discuten las tres definiciones de la probabilidad siguientes:

- La definición clásica.
- La definición de la probabilidad como frecuencia relativa.
- La definición axiomática.

Definición clásica. La probabilidad $\text{Prob}\{A\}$ de un evento A puede determinarse a priori, sin necesidad de experimentación, de la manera siguiente:

$$\text{Prob}\{A\} = \frac{N_A}{N} \quad (\text{A.1})$$

donde N es el número de posibles resultados del experimento y N_A es el número de estos resultados que producen el evento A .

Ejemplo A.2.1. En el experimento “lanzamiento de una moneda”:

$$\text{Casos posibles} = \{\text{cara}, \text{cruz}\} \quad (\text{A.2})$$

$$\text{Prob}\{\text{cara}\} = \frac{1}{2} \quad \text{Prob}\{\text{cruz}\} = \frac{1}{2} \quad (\text{A.3})$$

□

Ejemplo A.2.2. En el experimento “lanzamiento de un dado”:

$$\text{Casos posibles} = \{1, 2, 3, 4, 5, 6\} \quad (\text{A.4})$$

$$\text{Prob}\{1\} = \text{Prob}\{2\} = \dots = \text{Prob}\{6\} = \frac{1}{6} \quad (\text{A.5})$$

$$\text{Prob}\{\text{Impar}\} = \frac{3}{6} \quad (\text{A.6})$$

$$\text{Prob}\{\text{Mayor que } 1\} = \frac{5}{6} \quad (\text{A.7})$$

□

Sin embargo, esta definición sólo es válida cuando el número de casos posibles es finito y además *todos ellos son equiprobables*. Así, por ejemplo, sería ilógico estudiar la probabilidad de que ocurra un accidente de avión de la forma siguiente:

$$\text{Casos posibles} = \{\text{Accidente}, \text{No accidente}\} \quad (\text{A.8})$$

$$\text{Prob}\{\text{Accidente}\} = \frac{1}{2} \quad (\text{A.9})$$

Obsérvese que al afirmar que “la definición clásica de probabilidad es sólo aplicable cuando todos los eventos posibles son igualmente probables”, se está empleando en la definición el concepto definido. Sin embargo, la definición clásica fue introducida como consecuencia del *principio de razón insuficiente*:

”si se posee ningún conocimiento a priori, se debe asumir que los posibles resultados del experimento son equiprobables”.

Definición de la probabilidad como frecuencia relativa. La probabilidad de que suceda el evento A se define de la forma siguiente:

$$\text{Prob}\{A\} = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (\text{A.10})$$

donde n_A es el número de experimentos en que ocurre A y n es el número total de experimentos.

Sin embargo, puesto que un experimento físico no puede repetirse infinitas veces, en la práctica se aplica la definición (inherentemente imprecisa) dada al comienzo de la sección.

Definición axiomática de la probabilidad. Fue propuesta en 1933 por Kolmogoroff, y se basa únicamente en los siguientes tres postulados:

- La probabilidad $\text{Prob}\{A\}$ de un evento A es un número positivo asignado a ese evento: $\text{Prob}\{A\} \geq 0$.
- La probabilidad de un evento, E , que se obtiene como resultado de todas las repeticiones del experimento, es igual a uno: $\text{Prob}\{E\} = 1$.
- La unión $A + B$ de dos eventos, A y B , es un evento que se produce cuando sucede A , cuando sucede B o cuando suceden ambos simultáneamente.

Se dice que dos eventos son *mutuamente excluyentes* cuando el hecho de que ocurra uno impide que ocurra el otro. Si dos eventos A y B son mutuamente excluyentes, entonces:

$$\text{Prob}\{A + B\} = \text{Prob}\{A\} + \text{Prob}\{B\} \quad (\text{A.11})$$

A.3. Probabilidad condicionada

El concepto de probabilidad condicionada surge para responder a preguntas del tipo: se sabe que se ha producido el evento A_2 , ¿cuál es la probabilidad de que suceda A_1 ? Esta probabilidad de que suceda A_1 , “condicionada” al hecho de que suceda A_2 , se representa: $\text{Prob}\{A_1 | A_2\}$, y satisface:

$$\text{Prob}\{A_1 | A_2\} = \frac{\text{Prob}\{A_1 \cdot A_2\}}{\text{Prob}\{A_2\}} \rightarrow \text{Prob}\{A_1 \cdot A_2\} = \text{Prob}\{A_1 | A_2\} \cdot \text{Prob}\{A_2\} \quad (\text{A.12})$$

es decir, la probabilidad de que suceda tanto A_1 como A_2 , $\text{Prob}\{A_1 \cdot A_2\}$, es igual a la probabilidad de que suceda A_2 , $\text{Prob}\{A_2\}$, multiplicada por la probabilidad de que suceda A_1 habiendo sucedido A_2 , $\text{Prob}\{A_1 | A_2\}$.

Cuando los eventos A_1 y A_2 son *independientes* (es decir, que ocurra uno de los dos eventos no afecta a la probabilidad de que ocurra el otro), entonces:

$$\text{Prob}\{A_1 | A_2\} = \text{Prob}\{A_1\} \quad (\text{A.13})$$

El concepto de independencia estadística entre eventos es un concepto clave, sobre el que se basa el desarrollo matemático de la teoría de la probabilidad. Como regla general, *la independencia física es una indicación suficiente de la independencia estadística*. Por consiguiente, se puede considerar que aquellos eventos que no tienen ninguna conexión física entre sí son estadísticamente independientes.

A.4. El teorema fundamental de la probabilidad

Sea un experimento con dos posibles resultados:

{Se produce el evento A , No se produce el evento A }

y sea p la probabilidad de que se produzca el evento A .

El teorema establece que si se repite el experimento n veces, entonces la probabilidad de que el suceso A ocurra n_A veces (en cualquier orden) es:

$$\text{Prob}\{A \text{ ocurre } k \text{ veces en cualquier orden}\} = \binom{n}{n_A} \cdot p^{n_A} \cdot (1-p)^{n-n_A} \quad (\text{A.14})$$

donde:

$$\binom{n}{n_A} = \frac{n!}{n_A! \cdot (n - n_A)!} \quad (\text{A.15})$$

La aplicación práctica de la fórmula $\binom{n}{n_A} \cdot p^{n_A} \cdot (1 - p)^{n - n_A}$ es muy problemática: la función factorial crece muy rápidamente y enseguida desborda el máximo número representable en la calculadora. Por este motivo, se han desarrollado aproximaciones a esta fórmula que permitan su aplicación práctica. Dos de ellas son las siguientes:

Teorema de DeMoivre-Laplace. Si $n \cdot p \cdot (1 - p) \gg 1$, entonces:

$$\binom{n}{n_A} \cdot p^{n_A} \cdot (1 - p)^{n - n_A} \approx \frac{1}{\sqrt{2 \cdot \pi \cdot n \cdot p \cdot (1 - p)}} \cdot e^{-\frac{(n_A - n \cdot p)^2}{2 \cdot n \cdot p \cdot (1 - p)}} \quad (\text{A.16})$$

Es decir, puede aproximarse por la densidad de probabilidad de una distribución normal con $x = n_A$, media $\mu = n \cdot p$ y varianza $\sigma^2 = n \cdot p \cdot (1 - p)$. Esta aproximación es una igualdad en el límite $n \rightarrow \infty$.

Teorema de Poisson. Si $p \ll 1$, y además n_A es del orden de $n \cdot p$, entonces:

$$\binom{n}{n_A} \cdot p^{n_A} \cdot (1 - p)^{n - n_A} \approx e^{-n \cdot p} \cdot \frac{(n \cdot p)^{n_A}}{n_A!} \quad (\text{A.17})$$

La formulación de este teorema en el límite es la siguiente:

$$\text{Si: } \left\{ \begin{array}{l} n \rightarrow \infty \\ p \rightarrow 0 \\ n \cdot p \rightarrow a \end{array} \right\} \text{ entonces: } \binom{n}{n_A} \cdot p^{n_A} \cdot (1 - p)^{n - n_A} \rightarrow e^{-a} \cdot \frac{a^{n_A}}{n_A!} \quad (\text{A.18})$$

A.5. El concepto de variable aleatoria

El *espacio muestral*, S , de un experimento aleatorio es el conjunto de todos los resultados posibles del experimento. Los elementos del espacio deben excluirse mutuamente, es decir, sólo uno de ellos puede ser el resultado del experimento cada vez.

Una *variable aleatoria*, X , es una función que asocia un número a cada uno de los elementos del espacio muestral, S . Esta relación está representada en la Figura A.1.

En este texto se emplea el criterio de emplear letras mayúsculas para designar a las variables aleatorias y letras minúsculas para representar las observaciones (o realizaciones) de las variables aleatorias. Por ejemplo, una observación de la variable aleatoria X se representaría: x .

Obsérvese que, dado que una variable aleatoria es una función, no es correcto hablar de “generar una variable aleatoria”, sino que debe hablarse de “generar *observaciones* (o realizaciones) de una variable aleatoria”. La generación de observaciones se realiza “*muestreando*” la distribución de probabilidad de la variable aleatoria.

Las variables aleatorias se clasifican en discretas y continuas:

- Una variable aleatoria se dice *discreta* cuando su espacio muestra tiene un conjunto numerable de elementos, es decir, cuando sólo puede tomar un conjunto de valores finito o bien un conjunto de valores infinito numerable.
- Una variable aleatoria es *continua* cuando su espacio muestra es un conjunto no numerable de elementos. Las variables aleatorias continuas son una idealización matemática, ya que cualquier procedimiento de medida que se emplee tendrá un límite de precisión. Igualmente, al programar cualquier algoritmo de generación de números aleatorios éstos se obtendrán con un determinado número finito de cifras decimales. Con lo cual, en la práctica el espacio muestra será numerable.

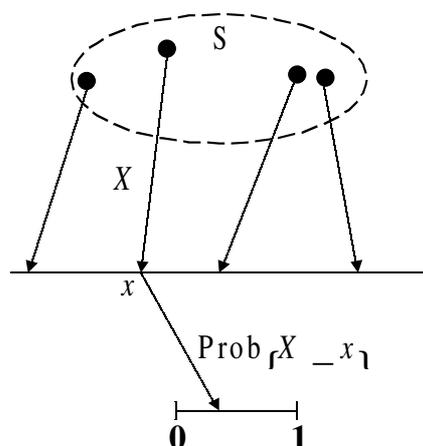


Figura A.1: Espacio muestral, variable aleatoria y probabilidad.

A.6. Probabilidad acumulada y densidad de probabilidad

Probabilidad acumulada. El número de elementos del espacio muestral, S , que son englobados por el evento $\{X \leq x\}$ va aumentando al aumentar el número x . Por tanto, la probabilidad de dicho evento, $\text{Prob}\{X \leq x\}$, depende por tanto del valor de x . Esta probabilidad se representa $F_X(x)$ y se denomina *función de probabilidad acumulada* de la variable aleatoria X :

$$F_X(x) = \text{Prob}\{X \leq x\} \tag{A.19}$$

Densidad de probabilidad. La derivada

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{A.20}$$

se denomina *función densidad de probabilidad* de la variable aleatoria X .

A.7. Cuantiles y mediana

Cuantiles. Se denomina *percentil* o *cuantil* q de una variable aleatoria X , al número x_q más pequeño que satisface:

$$q = \text{Prob}\{X \leq x_q\} = F_X(x_q) \tag{A.21}$$

Así pues, los cuantiles de una variable aleatoria se calculan de la inversa de la probabilidad acumulada, de la forma siguiente:

$$x_q = F_X^{-1}(q) \tag{A.22}$$

El dominio de la función inversa, F_X^{-1} , es el intervalo $0 \leq q \leq 1$ y su rango el eje real.

Mediana. La mediana de una variable aleatoria X es el número más pequeño, m , que satisface:

$$F_X(m) = \text{Prob}\{X \leq m\} = 0.5 \tag{A.23}$$

Así pues, la mediana es el cuantil 0.5.

A.8. Media

Media o valor esperado. La media o valor esperado de una variable aleatoria X se representa indistintamente $E\{X\}$ o μ_X , y se define de la forma siguiente:

$$E\{X\} = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx \quad (\text{A.24})$$

Si la variable aleatoria X es de tipo discreto, y toma valores $\{x_1, x_2, \dots\}$, con probabilidad $\{p_X(x_1), p_X(x_2), \dots\}$, entonces la media puede expresarse de la forma:

$$E\{X\} = \sum_i x_i \cdot p_X(x_i) \quad \text{donde } p_X(x_i) = \text{Prob}\{X = x_i\} \quad (\text{A.25})$$

Media de $g(X)$. Dada una variable aleatoria X y una función $g(x)$, la media de la variable aleatoria

$$Y = g(X) \quad (\text{A.26})$$

se calcula de la forma siguiente:

$$E\{Y\} = \int_{-\infty}^{\infty} y \cdot f_Y(y) \cdot dy = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \cdot dx \quad (\text{A.27})$$

Si X es de tipo discreto:

$$E\{Y\} = \sum_j y_j \cdot p_Y(y_j) = \sum_i g(x_i) \cdot p_X(x_i) \quad (\text{A.28})$$

A.9. Varianza y desviación estándar

Varianza. La varianza de una variable aleatoria X se representa σ_X^2 o $\text{var}\{X\}$, y se define de la forma siguiente¹:

$$\sigma_X^2 = E\{(x - E\{X\})^2\} = E\{x^2\} - (E\{X\})^2 \quad (\text{A.29})$$

En el caso continuo, se calcula de la forma siguiente:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - E\{X\})^2 \cdot f_X(x) \cdot dx \quad (\text{A.30})$$

En el caso discreto:

¹Demostración:

$$\begin{aligned} \sigma_X^2 &= E\{(x - E\{X\})^2\} \\ &= E\{x^2 + (E\{X\})^2 - 2 \cdot x \cdot E\{X\}\} \\ &= E\{x^2\} + (E\{X\})^2 - 2 \cdot E\{X\} \cdot E\{X\} \\ &= E\{x^2\} - (E\{X\})^2 \quad \square \end{aligned}$$

$$\sigma_X^2 = \sum_i (x_i - E\{X\})^2 \cdot p_X(x_i) \quad (\text{A.31})$$

Desviación estándar. La desviación estándar, σ_X , de una variable aleatoria X , se define como la raíz cuadrada de la varianza:

$$\sigma_X = \sqrt{\sigma_X^2} \quad (\text{A.32})$$

A.10. Estimadores de la media y de la varianza

Dado un conjunto de n observaciones independientes, x_1, \dots, x_n , de una variable aleatoria X , se definen los dos siguientes estadísticos:

$$\bar{X}(n) = \frac{\sum_{i:1}^n x_i}{n} \quad (\text{Media aritmética}) \quad (\text{A.33})$$

$$S_X^2(n) = \frac{\sum_{i:1}^n (x_i - \bar{X}(n))^2}{n-1} \quad (\text{A.34})$$

Estos dos estadísticos, que a su vez son variables aleatorias, son estimadores no sesgados de la media y la varianza de X respectivamente:

$$E\{\bar{X}(n)\} = \mu_X \quad (\text{A.35})$$

$$E\{S_X^2(n)\} = \sigma_X^2 \quad (\text{A.36})$$

La varianza de los estimadores, que proporciona una medida de su dispersión respecto a su media, puede estimarse de la forma siguiente:

$$\hat{\sigma}_{\bar{X}(n)}^2 = \frac{S_X^2(n)}{n} \quad (\text{A.37})$$

$$\text{var}\{S_X^2(n)\} = \frac{2 \cdot (S_X^2(n))^2}{n-1} \quad (\text{A.38})$$

Se observa que al aumentar el tamaño de la muestra, n , disminuye la varianza de los estimadores (ver Figura A.2). En el límite, al tender el tamaño de muestra a infinito, la varianza de ambos estimadores tiende a cero, es decir, aumenta la probabilidad de que la muestra obtenida del estimador esté próxima al “valor verdadero” que se pretende estimar:

$$\begin{aligned} \bar{X}(n) &\rightarrow \mu_X \\ S^2(n) &\rightarrow \sigma_X^2 \end{aligned} \quad (\text{A.39})$$

A.11. Covarianza y coeficiente de correlación

Covarianza. La covarianza de dos variables aleatorias X e Y se define de la forma siguiente:

$$\begin{aligned} \text{cov}(X, Y) &= E\{(X - \mu_X) \cdot (Y - \mu_Y)\} \\ &= E\{X \cdot Y\} - \mu_X \cdot \mu_Y \end{aligned} \quad (\text{A.40})$$

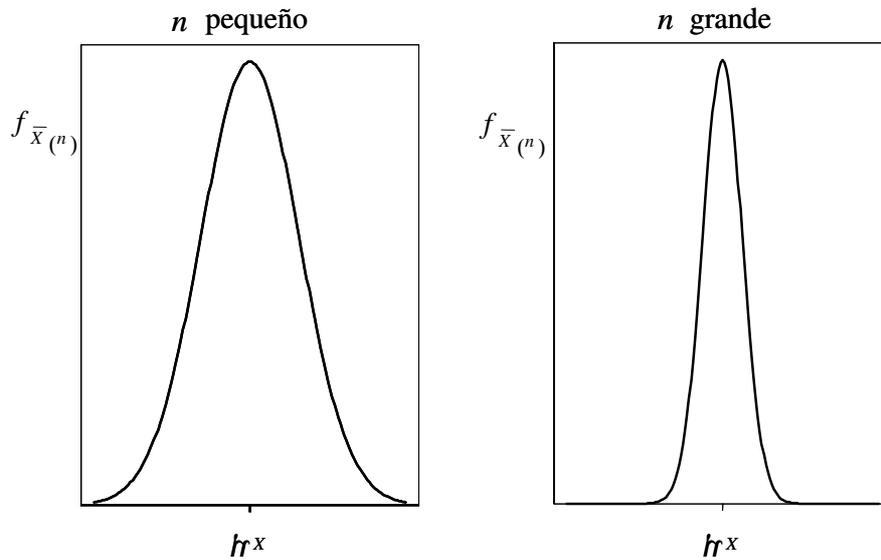


Figura A.2: La varianza de la distribución del estimador de la media disminuye al aumentar el tamaño de la muestra.

La *covarianza* de X e Y es una medida de la dependencia lineal entre ambas variables aleatorias. Su interpretación de la *covarianza* es muy intuitiva:

- Cuando $\text{cov}(X, Y) > 0$, significa que tiende a ocurrir uno de los siguientes dos eventos o ambos:

$$\{X > \mu_X, Y > \mu_Y\} \quad \{X < \mu_X, Y < \mu_Y\} \quad (\text{A.41})$$

Es decir, cuando una de las dos variables toma un valor grande la otra también tiende a ser grande. Por el contrario, cuando una de las variables toma un valor pequeño, la otra tiende a ser pequeña. Se dice que las variables están *correlacionadas positivamente*.

- Cuando $\text{cov}(X, Y) < 0$, significa que tiende a ocurrir uno de los siguientes dos eventos o ambos:

$$\{X > \mu_X, Y < \mu_Y\} \quad \{X < \mu_X, Y > \mu_Y\} \quad (\text{A.42})$$

Es decir, cuando una variable toma un valor grande, la otra tiende a tomar un valor pequeño y viceversa. Se dice que ambas variables están *correlacionadas negativamente*.

- Cuando $\text{cov}(X, Y) = 0$, se dice que las dos variables *no están correlacionadas*. Si dos variables son independientes, entonces no están correlacionadas. Sin embargo, el hecho de que dos variables no estén correlacionadas no implica que sean independientes.

Coefficiente de correlación. La covarianza depende de las unidades en que están expresadas las variables aleatorias, con lo cual su interpretación puede resultar confusa. Para evitar esta dificultad, se define el *coeficiente de correlación*, entre dos variables aleatorias X e Y , de la forma siguiente:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (\text{A.43})$$

La interpretación del coeficiente de correlación resulta sencilla, ya que:

$$|\rho_{X,Y}| \leq 1 \quad (\text{A.44})$$

Dado que la desviación estándar siempre es un número positivo, el coeficiente de correlación y la covarianza tienen el mismo signo:

- Si $\rho_{X,Y}$ está próximo a $+1$, entonces X e Y están fuertemente correlacionadas positivamente.
- Si $\rho_{X,Y}$ está próximo a -1 , entonces las dos variables están fuertemente correlacionadas negativamente.
- Si $\rho_{X,Y} = 0$, entonces no están correlacionadas.

A.12. Secuencias de variables aleatorias

Proceso estocástico. Un proceso estocástico es una colección ordenada temporalmente de variables aleatorias, definidas todas sobre el mismo espacio muestral. Esta colección de variables aleatorias puede representarse de la forma siguiente:

$$\{X_t, t \in T\} \quad (\text{A.45})$$

donde T es el conjunto de todos los valores que puede tomar el índice t . El conjunto de valores del índice t puede ser de los dos tipos siguientes:

- Si T es contable (habitualmente los enteros positivos), se dice que se trata de un proceso estocástico de *tiempo discreto*.
- Si T es un subconjunto incontable de los números reales (normalmente los reales no negativos), entonces se dice que el proceso estocástico es de *tiempo continuo*.

Procesos estocásticos con covarianza estacionaria. A la hora de analizar los datos de salida de la simulación, se realizan hipótesis acerca del proceso estocástico al que corresponden los datos. Estas hipótesis son aproximaciones (sin las cuales el análisis estadístico de los datos no sería posible) que frecuentemente no son estrictamente ciertas en la práctica. Una de estas hipótesis es asumir que el proceso estacionario tiene *covarianza estacionaria*.

Un proceso estocástico de tiempo discreto, X_1, X_2, \dots , se dice que tiene *covarianza estacionaria* cuando:

- La media y la varianza de todas las variables del proceso es iguales, y éstas son finitas:

$$\begin{aligned} \mu_i &= \mu & \text{para } i = 1, 2, \dots \text{ y } -\infty < \mu < \infty \\ \sigma_i^2 &= \sigma^2 & \text{para } i = 1, 2, \dots \text{ y } \sigma^2 < \infty \end{aligned} \quad (\text{A.46})$$

- La covarianza entre dos variables aleatorias cualquiera del proceso, X_i y X_{i+j} , depende únicamente de la separación entre ellas, j , y no de su posición en la secuencia, i e $i+j$:

$$\text{cov}\{X_i, X_{i+j}\} \text{ es independiente de } i \text{ para } j : 1, 2, \dots \quad (\text{A.47})$$

A.13. Teorema del límite central

Supóngase que la variable aleatoria X distribuida arbitrariamente, con media μ_X y varianza σ_X^2 . Se realizan n observaciones, x_1, \dots, x_n , de la variable aleatoria. El teorema establece que:

Si el tamaño de la muestra, n , es lo suficientemente grande, entonces la variable aleatoria $\bar{X}(n)$ (es decir, la media aritmética de las observaciones) está distribuida

aproximadamente como una variable aleatoria normal con media μ_X y varianza $\frac{\sigma_X^2}{n}$:

$$N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \quad (\text{A.48})$$

Esta aproximación es una igualdad en el límite $n \rightarrow \infty$.

Obsérvese que el teorema no impone ninguna condición acerca de la distribución de X : se cumple con independencia de la forma en que esté distribuida X .

Asimismo, afirmar que

la distribución de $\bar{X}(n)$ tiende a $N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$ cuando $n \rightarrow \infty$

es equivalente a afirmar que

la distribución de $\frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}}$ tiende a $N(0,1)$ cuando $n \rightarrow \infty$

y también es equivalente (dado que $S_X^2(n)$ converge a σ_X^2 cuando $n \rightarrow \infty$), a afirmar que

la distribución de $\frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{S_X^2(n)}{n}}}$ tiende a $N(0,1)$ cuando $n \rightarrow \infty$

A.14. Intervalos de confianza para la media

Puntos críticos de la distribución normal estándar. El punto crítico $z_{1-\frac{\alpha}{2}}$ de una distribución normal estándar (media cero y varianza unidad) se define de la forma siguiente:

$$\text{Prob}\left\{-z_{1-\frac{\alpha}{2}} \leq X \leq z_{1-\frac{\alpha}{2}}\right\} = 1 - \alpha \quad \text{donde } X \sim N(0,1) \quad (\text{A.49})$$

En la última fila de la Tabla T.1, situada al final del libro, pueden encontrarse los puntos críticos de la distribución normal estándar.

Intervalo de confianza para la media. En virtud del teorema del límite central:

la distribución de $\frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{S_X^2(n)}{n}}}$ tiende a $N(0,1)$ cuando $n \rightarrow \infty$

En consecuencia, cuando n es grande:

$$\text{Prob}\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{S_X^2(n)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right\} \approx 1 - \alpha \quad (\text{A.50})$$

o, equivalentemente

$$\text{Prob}\left\{\bar{X}(n) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \leq \mu_X \leq \bar{X}(n) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}}\right\} \approx 1 - \alpha \quad (\text{A.51})$$

con lo cual, si n es grande,

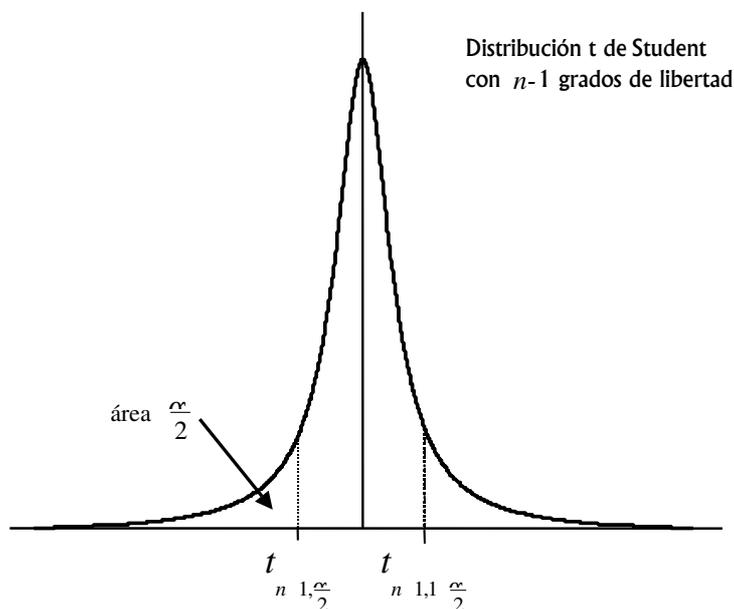


Figura A.3: Puntos críticos de la distribución t de Student.

$$\bar{X}(n) \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \tag{A.52}$$

es un intervalo de confianza del $100 \cdot (1 - \alpha) \%$ para la media, μ_X .

Una dificultad asociada a esta definición de los intervalos de confianza es determinar en cada caso qué significa que n sea “suficientemente grande”. Cuanto más sesgada (no simétrica) sea la distribución de X , mayor necesita ser n para que la distribución de $\frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{S_X^2(n)}{n}}}$ se aproxime a $N(0,1)$. Si n es demasiado pequeño, el intervalo obtenido será demasiado pequeño, es decir, la confianza del intervalo es en realidad menor que $100 \cdot (1 - \alpha) \%$.

Interpretación del intervalo de confianza. Para un determinado conjunto x_1, \dots, x_n de muestras de la variable aleatoria X , los extremos del intervalo para μ_X son números. Es decir, el intervalo calculado de la muestra o bien contiene a la media μ_X o no la contiene. No hay nada aleatorio acerca de un intervalo obtenido de un determinado conjunto de datos. La correcta interpretación de un intervalo de confianza es la siguiente:

Si se construye un número muy grande de intervalos independientes con una confianza del $100 \cdot (1 - \alpha) \%$, cada uno basado en n muestras independientes de X , donde n es suficientemente grande, la proporción de estos intervalos que contienen a μ_X es $1 - \alpha$.

Intervalo de confianza para la media, cuando X es normal. Si las observaciones x_1, \dots, x_n de X son independientes (con $n \geq 2$) y $X \sim N(\mu_X, \sigma_X^2)$, entonces la variable aleatoria

$$\frac{\bar{X}(n) - \mu_X}{\sqrt{\frac{S^2(n)}{n}}} \sim t \text{ de Student con } n - 1 \text{ grados de libertad}$$

y

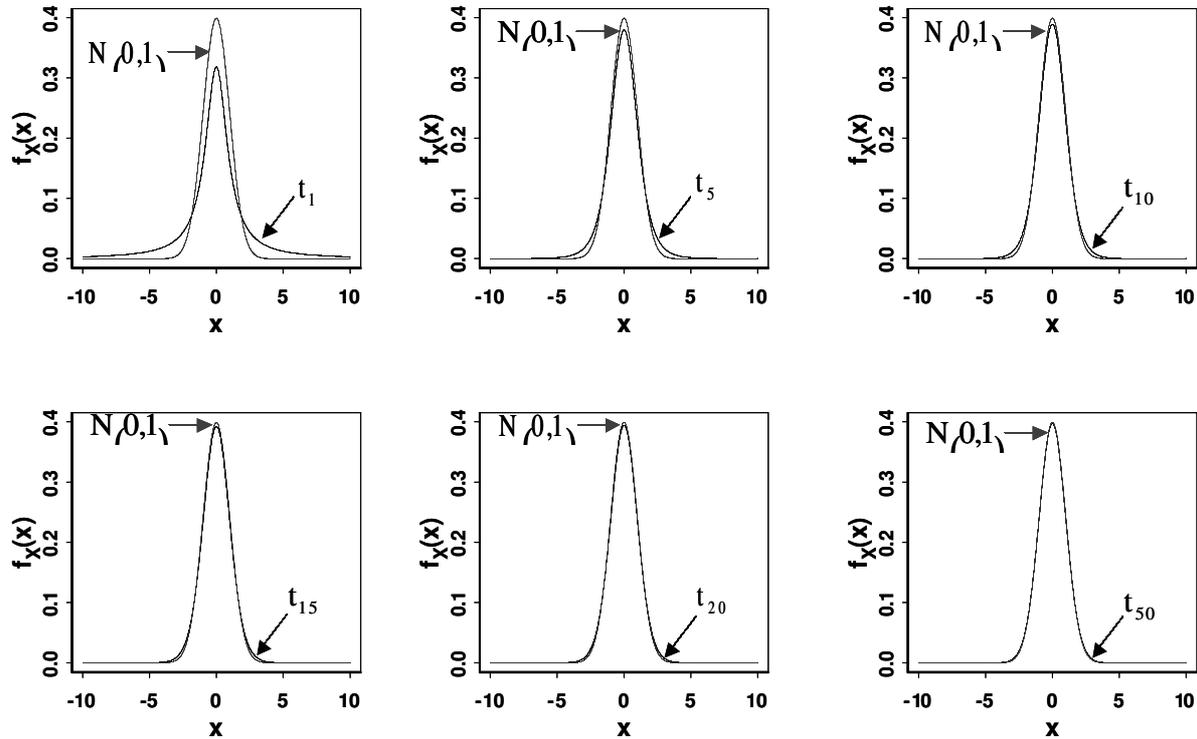


Figura A.4: Distribución normal y t de Student con 1, 5, 10, 15, 20 y 50 grados de libertad.

$$\bar{X}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \quad (\text{A.53})$$

es un intervalo de confianza *exacto* del $100 \cdot (1 - \alpha) \%$ para μ_X , donde $t_{n-1, 1-\frac{\alpha}{2}}$ es el punto crítico para la distribución t con $n - 1$ grados de libertad (ver Figura A.3). Es decir:

$$\text{Prob} \{T_{n-1} \leq t_{n-1, 1-\frac{\alpha}{2}}\} = 1 - \frac{\alpha}{2} \quad (\text{A.54})$$

La distribución t de Student es simétrica entorno al cero, $t_{n-1, \frac{\alpha}{2}} = -t_{n-1, 1-\frac{\alpha}{2}}$, y se aproxima a una normal estándar cuando el número de grados de libertad tiende a infinito (ver Figura A.4). En la Tabla T.1, situada al final de libro, pueden encontrarse los puntos críticos de la distribución t de Student.

Consideraciones prácticas. En la práctica, la distribución de la que se muestrean las observaciones x_1, \dots, x_n raramente es normal. Si no es normal, el intervalo de confianza representado por la Ecuación (A.53) no es exacto, sino aproximado.

Dado que la distribución t de Student es más baja y abierta que la distribución $N(0,1)$ (ver Figura A.4), los intervalos de confianza t siempre son más anchos que los de la distribución normal:

$$t_{n-1, 1-\frac{\alpha}{2}} > z_{1-\frac{\alpha}{2}} \quad (\text{A.55})$$

Al ser más anchos, frecuentemente los intervalos t tienen un recubrimiento más próximo al nivel de confianza deseado, por eso suele ser más recomendable su uso aun cuando las muestras no estén distribuidas normalmente. Por otra parte, como puede verse en la Figura

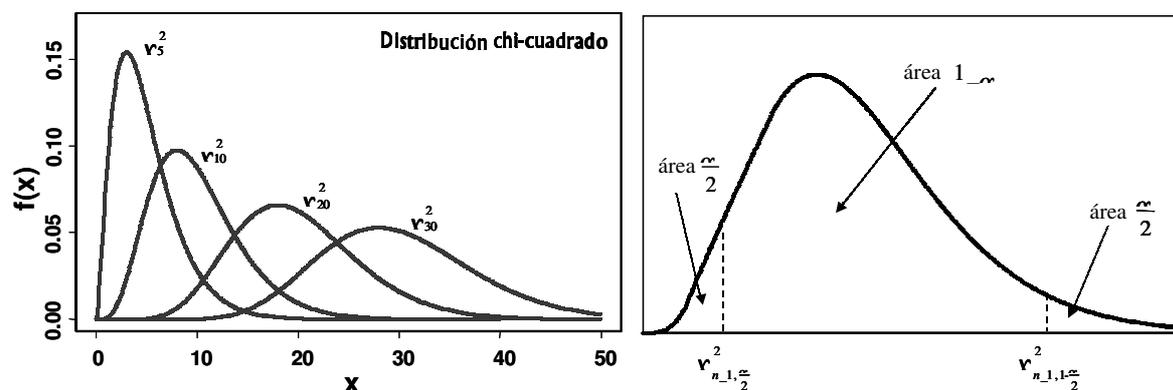


Figura A.5: Densidad de probabilidad de algunas distribuciones chi-cuadrado (izda). Puntos críticos de la distribución chi-cuadrado con $n - 1$ grados de libertad (drcha).

A.4, la diferencia entre la distribución t y la normal estándar es relevante sólo para valores de n pequeños.

Ejemplo A.14.1. Se han realizado las siguientes 10 observaciones independientes de una distribución normal con media y varianza desconocida.

| | | | | |
|-------|--------|-------|-------|--------|
| 0.202 | 0.498 | 0.680 | 0.888 | 0.486 |
| 0.089 | -0.048 | 0.583 | 0.553 | -0.497 |

Se desea construir un intervalo de confianza del 90% para la media de la distribución:

$$100 \cdot (1 - \alpha) = 90 \rightarrow \alpha = 0.1 \quad \bar{X}(10) = 0.343 \quad S_X^2(10) = 0.167$$

$$\bar{X}(10) \pm t_{9,0.95} \cdot \sqrt{\frac{S_X^2(10)}{10}} = 0.343 \pm 1.833 \cdot \sqrt{\frac{0.167}{10}} = 0.343 \pm 0.237$$

Así pues, puede afirmarse con un 90% de confianza que μ_X se encuentra en el intervalo $[0.106, 0.580]$. \square

A.15. Intervalos de confianza para la varianza de una distribución normal

Sea x_1, \dots, x_n un conjunto de observaciones independientes, obtenidas muestreando la distribución $N(\mu_X, \sigma_X^2)$. Entonces, la variable aleatoria $\frac{(n-1) \cdot S_X^2(n)}{\sigma_X^2}$ está distribuida chi-cuadrado con $n - 1$ grados de libertad (ver Figura A.5):

$$\frac{(n - 1) \cdot S_X^2(n)}{\sigma_X^2} \sim \chi_{n-1}^2 \tag{A.56}$$

Esta propiedad permite la construcción de intervalos de confianza para la varianza de las distribuciones normales. Para ello, se emplea de nuevo el concepto de punto crítico de la distribución. Se define el *punto crítico* $\chi_{n-1, \alpha}^2$ de la distribución chi-cuadrado con $n - 1$ grados de libertad de la forma siguiente (ver Figura A.5):

$$\text{Prob} \{ \chi_{n-1}^2 \leq \chi_{n-1, \alpha}^2 \} = \alpha \quad (\text{A.57})$$

En la Tabla T.2, situada al final del libro, se muestran los puntos críticos de la distribución chi-cuadrado.

De lo anterior, se deduce:

$$\text{Prob} \left\{ \chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{(n-1) \cdot S_X^2(n)}{\sigma_X^2} \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right\} = 1 - \alpha \quad (\text{A.58})$$

o, equivalentemente, se obtiene la expresión del intervalo de confianza para la varianza:

$$\text{Prob} \left\{ \frac{(n-1) \cdot S_X^2(n)}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma_X^2 \leq \frac{(n-1) \cdot S_X^2(n)}{\chi_{n-1, \frac{\alpha}{2}}^2} \right\} = 1 - \alpha \quad (\text{A.59})$$

Ejemplo A.15.1. *El procedimiento que sigue un laboratorio, para determinar la concentración de una sustancia tóxica disuelta en el agua, consiste en toma 16 muestras de agua y analizarlas. Los resultados del análisis son los siguientes (en mg/litro):*

| | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 5.176 | 5.075 | 4.694 | 4.889 | 4.920 | 4.929 | 4.800 | 5.216 |
| 5.157 | 4.986 | 4.790 | 4.810 | 5.207 | 4.996 | 5.166 | 4.972 |

La máxima concentración permitida de la sustancia tóxica es 5.1 mg/litro. ¿Se puede afirmar con una confianza del 95% que este proceso cumple la normativa? También se desea calcular un intervalo de confianza del 95% para la desviación típica.

Los estimadores de la media y la desviación estándar son los siguientes:

$$\bar{X}(16) = 4.986 \text{ mg/litro} \quad S_X(16) = 0.166 \text{ mg/litro} \quad (\text{A.60})$$

Los intervalos de confianza del 95% para la media y la varianza son:

$$\bar{X}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} = 4.986 \pm 2.131 \cdot \sqrt{\frac{0.166^2}{16}} = 4.986 \pm 0.088 \quad (\text{A.61})$$

$$\frac{(n-1) \cdot S_X^2(n)}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma_X^2 \leq \frac{(n-1) \cdot S_X^2(n)}{\chi_{n-1, \frac{\alpha}{2}}^2} \rightarrow \frac{15 \cdot 0.166^2}{27.488} \leq \sigma_X^2 \leq \frac{15 \cdot 0.166^2}{6.26} \quad (\text{A.62})$$

En consecuencia:

$$4.898 \leq \mu_X \leq 5.074 \quad 0.015 \leq \sigma_X^2 \leq 0.066 \quad (\text{A.63})$$

Dado que el intervalo para la media no incluye 5.1 mg/litro, se puede afirmar con una confianza del 95% que el proceso cumple la normativa. \square

A.16. Intervalo de confianza para la proporción de una distribución binomial

Se llama *experimento de Bernoulli* aquel que tiene dos posibles resultados: la ocurrencia o no de cierto evento. La probabilidad p , con $p \in (0, 1)$, de que en un experimento suceda el evento, se denomina *proporción*.

Una variable aleatoria discreta, X , que representa el número de eventos que se producen (en cualquier orden) en n experimentos independientes de Bernoulli, está distribuida binomialmente:

$$X \sim \text{bin}(n, p) \quad (\text{A.64})$$

La probabilidad de obtener un número x de eventos en n experimentos es:

$$p_X(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{si } x \in \{0, 1, \dots, n\} \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (\text{A.65})$$

donde:

$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!} \quad (\text{A.66})$$

De acuerdo con el teorema de DeMoivre-Laplace, cuando

$$n \cdot p \cdot (1-p) \gg 1 \quad (\text{A.67})$$

la probabilidad de que el evento se produzca x veces en n experimentos puede escribirse, aproximadamente de la forma siguiente:

$$p_X(x) \approx \frac{1}{\sqrt{2 \cdot \pi \cdot n \cdot p \cdot (1-p)}} \cdot e^{-\frac{(x-n \cdot p)^2}{2 \cdot n \cdot p \cdot (1-p)}} \quad \text{si } x \in \{0, 1, \dots, n\} \quad (\text{A.68})$$

Es decir, puede evaluarse a partir de la densidad de probabilidad de una distribución normal, con media $\mu = n \cdot p$ y varianza $\sigma^2 = n \cdot p \cdot (1-p)$.

Con el fin de construir el intervalo de confianza para la proporción, resulta conveniente expresar la probabilidad en función, no del número de veces x que sucede el evento, sino de la proporción de veces sucede: $\frac{x}{n}$. A su vez, dado un conjunto de réplicas del experimento, este cociente es un estimador de la proporción:

$$\hat{p} = \frac{x}{n} \quad (\text{A.69})$$

Realizando el cambio de variable, se obtiene la probabilidad en función de \hat{p} :

$$p_X(\hat{p}) = \begin{cases} \binom{n}{n \cdot \hat{p}} \cdot p^{n \cdot \hat{p}} \cdot (1-p)^{n \cdot (1-\hat{p})} & \text{si } \hat{p} \in \{0, \frac{1}{n}, \dots, 1\} \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (\text{A.70})$$

de donde:

$$p_X(\hat{p}) \approx \frac{1}{n} \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \frac{p \cdot (1-p)}{n}}} \cdot e^{-\frac{(\hat{p}-p)^2}{2 \cdot \frac{p \cdot (1-p)}{n}}} \quad \text{si } \hat{p} \in \left\{0, \frac{1}{n}, \dots, 1\right\} \quad (\text{A.71})$$

que es el producto de $\frac{1}{n}$ por la densidad de probabilidad de una distribución normal de media p y varianza $\frac{p \cdot (1-p)}{n}$. En la Figura A.6 se muestra el valor exacto (barra) de la probabilidad y el valor aproximado (curva normal correspondiente) para diferentes valores de la proporción y del número de experimentos.

De la aproximación de la probabilidad por una normal, se obtiene el siguiente intervalo de confianza aproximado para la proporción:

$$\text{Prob} \left\{ \hat{p} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \right\} = 1 - \alpha \quad (\text{A.72})$$

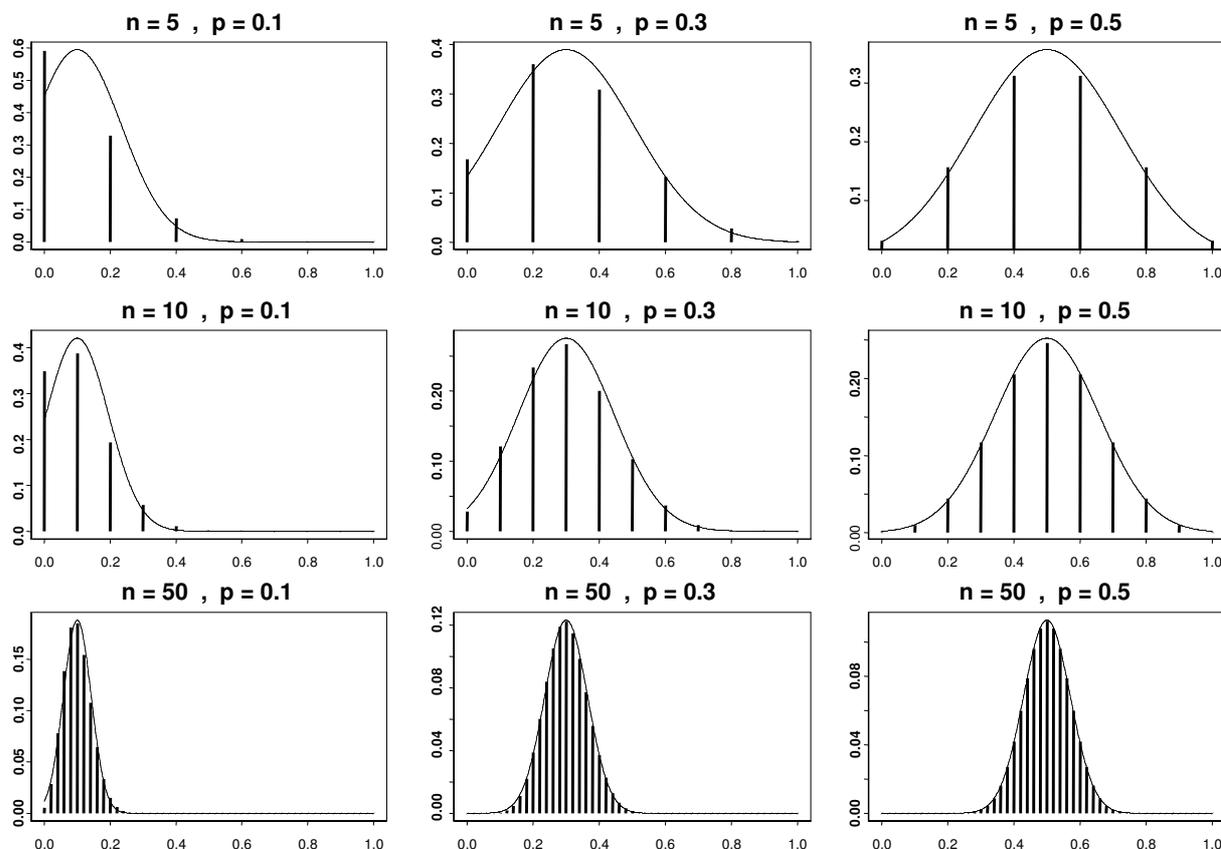


Figura A.6: Distribución bin (n, p) (barras) y su aproximación por una distribución normal.

A.17. Contraste de hipótesis paramétricas

¿Qué es? Una *hipótesis estadística paramétrica* es una suposición acerca del valor de uno o más parámetros de un modelo estadístico. Por ejemplo, que la media de una distribución tenga un determinado valor, que dos distribuciones tengan la misma varianza, etc. El *contraste de la hipótesis* es el proceso para establecer la validez de la misma, y se realiza a partir de una muestra x_1, \dots, x_n de la distribución.

En general, se sigue el convenio de llamar *hipótesis nula* y representarla H_0 , a aquella que se pretende de contrastar. La hipótesis alternativa se representa: H_1 .

Ejemplo A.17.1. Supóngase una variable aleatoria X , distribuida $X \sim F_X(x, \theta)$, donde θ es un parámetro de la distribución. Se desea contrastar la hipótesis $\theta = \theta_0$ frente a la alternativa, $\theta \neq \theta_0$, donde θ_0 es una constante. Se representa:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned} \tag{A.73}$$

□

El propósito del contraste de la hipótesis no es determinar si H_0 o H_1 son ciertas, sino *establecer si existe evidencia suficiente para rechazar H_0* . Los términos “aceptar” y “rechazar” deben ser interpretados en este contexto. Así, *no rechazar* la hipótesis nula no significa haber demostrado que es cierta, sino significa que *no hay evidencia suficiente en la muestra para afirmar que es falsa*.

Tipos de error del contraste. Pueden cometerse los dos tipos de errores siguientes:

- *Error de tipo I*, que consiste en rechazar H_0 cuando es verdadera.
- *Error de tipo II*, que es no rechazar H_0 cuando es falsa. Supone rechazar la hipótesis complementaria, H_1 , siendo ésta verdadera.

La probabilidad de cometer cada uno de estos errores es:

$$\alpha = \text{Prob} \{ \text{Error tipo I} \} = \text{Prob} \{ \text{Rechazar } H_0 | H_0 \text{ verdadera} \} \quad (\text{A.74})$$

$$\beta = \text{Prob} \{ \text{Error tipo II} \} = \text{Prob} \{ \text{No rechazar } H_0 | H_0 \text{ falsa} \} \quad (\text{A.75})$$

donde:

α se llama *nivel de significación del test*, y lo fija el experimentador que aplica el test.

β se llama *característica de operación* del test. Depende de H_1 y suele ser desconocido.

Existe una relación entre ambos errores: para un tamaño de muestra fijo, n , la reducción de α implica un aumento de β . Una forma de reducir β consiste en aumentar el tamaño de la muestra.

Se denomina *potencia del test* a la probabilidad de rechazar H_0 cuando es falsa:

$$P = 1 - \beta = \text{Prob} \{ \text{Rechazar } H_0 | H_0 \text{ falsa} \} \quad (\text{A.76})$$

Para α y H_1 fijas, la potencia del test sólo puede aumentarse aumentando el tamaño de la muestra, n . Dado que la potencia del test generalmente es desconocida y puede ser pequeña, prudentemente suele decirse que el test “no rechaza” (en vez de “acepta”) la hipótesis nula. Cuando la hipótesis H_0 no es rechazada por el test, generalmente no se sabe con qué probabilidad se puede afirmar que H_0 es cierta.

A.18. Contraste de hipótesis respecto a la media de una distribución normal

Se desea contrastar la hipótesis de que la media de cierta distribución normal vale un determinado valor fijo, μ_0 :

$$\begin{aligned} H_0 &: \mu_X = \mu_0 \\ H_1 &: \mu_X \neq \mu_0 \end{aligned} \quad (\text{A.77})$$

Para ello, se toman n muestras de la distribución: x_1, \dots, x_n . Si la hipótesis H_0 es cierta, entonces:

$$\frac{\bar{X}(n) - \mu_0}{\sqrt{\frac{S_X^2(n)}{n}}} \quad (\text{A.78})$$

está distribuida t de Student con $n - 1$ grados de libertad. Además,

$$\bar{X}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_X^2(n)}{n}} \quad (\text{A.79})$$

es un intervalo de confianza del $100 \cdot (1 - \alpha) \%$ para μ_0 .

El procedimiento para contrastar la hipótesis es el siguiente:

1. A partir de la muestra, x_1, \dots, x_n , construir un intervalo de confianza del $100 \cdot (1 - \alpha) \%$ para la media.
2. Examinar si μ_0 se encuentra dentro del intervalo. Si está dentro del intervalo, entonces no hay evidencia para rechazar la hipótesis nula. Si está fuera, se rechaza la hipótesis nula.

$$\left| \frac{\bar{X}(n) - \mu_0}{\sqrt{\frac{S_X^2(n)}{n}}} \right| \begin{cases} > t_{n-1, 1-\frac{\alpha}{2}} & \rightarrow \text{se rechaza } H_0 \\ \leq t_{n-1, 1-\frac{\alpha}{2}} & \rightarrow \text{no se rechaza } H_0 \end{cases} \quad (\text{A.80})$$

Ejemplo A.18.1. Se ha tomado la siguiente muestra de una distribución normal de media y varianza desconocidas:

$$\begin{array}{ccccc} 0.202 & 0.498 & 0.680 & 0.888 & 0.486 \\ 0.089 & -0.048 & 0.583 & 0.553 & -0.497 \end{array}$$

Se desea contrastar, con un nivel $\alpha = 0.1$, la hipótesis siguiente:

$$H_0 : \mu = 0 \quad (\text{A.81})$$

$$H_1 : \mu \neq 0 \quad (\text{A.82})$$

Para ello, se realiza el siguiente cálculo:

$$\frac{\bar{X}(n) - \mu_0}{\sqrt{\frac{S_X^2(n)}{n}}} = \frac{0.343 - 0}{\sqrt{\frac{0.167}{10}}} = 2.654 \quad (\text{A.83})$$

Dado que $t_{9, 0.95} = 1.833$, el test rechaza la hipótesis. □

A.19. Contraste de hipótesis respecto a la varianza de una distribución normal

Se desea contrastar la siguiente hipótesis respecto a la varianza de una distribución normal:

$$H_0 : \sigma_X^2 = \sigma_0^2 \quad (\text{A.84})$$

$$H_1 : \sigma_X^2 \neq \sigma_0^2 \quad (\text{A.85})$$

Para ello se toma una muestra de n observaciones independientes de la distribución, x_1, \dots, x_n . Si H_0 es cierta, entonces:

$$\frac{(n-1) \cdot S_X^2(n)}{\sigma_0^2} \sim \chi_{n-1}^2 \quad (\text{Distribución chi-cuadrado con } n-1 \text{ grados de libertad}) \quad (\text{A.86})$$

Con lo cual:

$$\text{Prob} \left\{ \chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{(n-1) \cdot S_X^2(n)}{\sigma_0^2} \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right\} = 1 - \alpha \quad (\text{A.87})$$

En consecuencia, el contraste de la hipótesis se realiza de la manera siguiente:

$$\frac{(n-1) \cdot S^2(n)}{\sigma_0^2} \begin{cases} \in \left[\chi_{n-1, \frac{\alpha}{2}}^2, \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right] & \rightarrow \text{no se rechaza } H_0 \\ \notin \left[\chi_{n-1, \frac{\alpha}{2}}^2, \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right] & \rightarrow \text{se rechaza } H_0 \end{cases} \quad (\text{A.88})$$

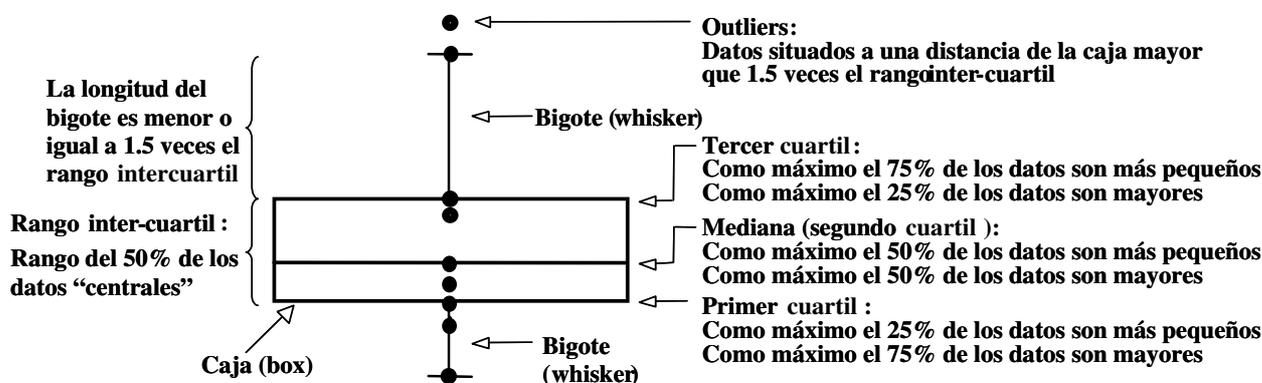


Figura A.7: Boxplot.

A.20. Contraste de hipótesis respecto a la proporción de una distribución binomial

Se desea contrastar la siguiente hipótesis acerca de la proporción de una distribución binomial:

$$H_0 : p = p_0 \tag{A.89}$$

$$H_1 : p \neq p_0 \tag{A.90}$$

Para ello, se realizan n réplicas independientes del experimento, obteniéndose que el evento sucede x veces. Se calcula la siguiente estimación de la proporción:

$$\hat{p} = \frac{x}{n} \tag{A.91}$$

En aquellas situaciones en que el teorema de DeMoivre-Laplace sea aplicable, puede contrastarse la hipótesis nula de la manera siguiente:

$$\left| \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot (1-p_0)}{n}}} \right| \begin{cases} \leq z_{1-\frac{\alpha}{2}} & \rightarrow \text{no hay evidencia para rechazar } H_0 \\ > z_{1-\frac{\alpha}{2}} & \rightarrow \text{se rechaza } H_0 \end{cases} \tag{A.92}$$

A.21. Boxplot (Box & Whiskers)

Los *boxplots* (también llamados *box & whiskers*) son una herramienta de análisis gráfico muy útil cuando se desea comparar la distribución de diferentes grupos de datos. En la Figura A.7 se indica cómo se construye el boxplot a partir de un conjunto de datos (representados mediante círculos).

Apéndice B

Distribuciones de probabilidad continuas

| UNIFORME | $U(a, b)$ |
|-----------------------------------|---|
| Densidad de probabilidad | $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$ |
| Parámetros | Dos parámetros reales, a y b , con $a < b$. Parámetro de posición: a . Parámetro de escala: $(b - a)$. |
| Rango | $[a, b]$ |
| Media | $\frac{a+b}{2}$ |
| Varianza | $\frac{(b-a)^2}{12}$ |
| Modo | No es único |
| Estimador de máxima verosimilitud | $\hat{a} = \min_{1 \leq i \leq n} \{x_i\}$ $\hat{b} = \max_{1 \leq i \leq n} \{x_i\}$ |

| | |
|-----------------------------------|---|
| EXPONENCIAL | expo (β) |
| Densidad de probabilidad | $f_X(x) = \begin{cases} \frac{1}{\beta} \cdot e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ |
| Parámetro | Parámetro de escala: $\beta > 0$ |
| Rango | $[0, \infty)$ |
| Media | β |
| Varianza | β^2 |
| Modo | 0 |
| Estimador de máxima verosimilitud | $\hat{\beta} = \bar{X}(n)$ |

GAMMA

gamma (α, β)

Densidad de probabilidad

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}}{\beta^\alpha \cdot \Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

donde $\Gamma(\alpha)$ es la función gamma.

Probabilidad acumulada

Si α no es un entero, no existe expresión analítica. Si α es un entero positivo, entonces:

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} \cdot \sum_{j:0}^{\alpha-1} \frac{(\frac{x}{\beta})^j}{j!} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Parámetro

Parámetro de forma: $\alpha > 0$.
Parámetro de escala: $\beta > 0$.

Rango

$$[0, \infty)$$

Media

$$\alpha \cdot \beta$$

Varianza

$$\alpha \cdot \beta^2$$

Modo

$$\begin{cases} \beta \cdot (\alpha - 1) & \text{si } \alpha \geq 1 \\ 0 & \text{si } \alpha < 1 \end{cases}$$

Estimador

$$\alpha = \left(\frac{\mu_X}{\sigma_X} \right)^2$$

$$\beta = \frac{\sigma_X^2}{\mu_X}$$

Para estimar el valor de los parámetros hay que sustituir la media y la varianza por sus estimadores: $\bar{X}_1(n)$ y $S_{\bar{X}}^2(n)$ respectivamente.

| WEIBULL | Weibull (α, β) |
|-----------------------------------|--|
| Densidad de probabilidad | $f_X(x) = \begin{cases} \alpha \cdot \beta^{-\alpha} \cdot x^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)^\alpha} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ |
| Parámetros | Parámetro de forma: $\alpha > 0$. Parámetro de escala: $\beta > 0$. |
| Rango | $[0, \infty)$ |
| Media | $\frac{\beta}{\alpha} \cdot \Gamma\left(\frac{1}{\alpha}\right)$ |
| Varianza | $\frac{\beta^2}{\alpha} \cdot \left[2 \cdot \Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \cdot \left(\Gamma\left(\frac{1}{\alpha}\right)\right)^2 \right]$ |
| Modo | $\beta \cdot \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} \quad \text{si } \alpha \geq 1$ $0 \quad \text{si } \alpha < 1$ |
| Estimador de máxima verosimilitud | Debe resolverse $\hat{\alpha}$ de la ecuación: $\frac{\sum_{i=1}^n x_i^{\hat{\alpha}} \cdot \ln(x_i)}{\sum_{i=1}^n x_i^{\hat{\alpha}}} - \frac{1}{\hat{\alpha}} = \frac{\sum_{i=1}^n \ln(x_i)}{n}$ Una vez calculado $\hat{\alpha}$, se calcula $\hat{\beta}$: $\hat{\beta} = \left(\frac{\sum_{i=1}^n x_i^{\hat{\alpha}}}{n} \right)^{\frac{1}{\hat{\alpha}}}$ |

| | |
|--------------------------|--|
| NORMAL | $N(\mu, \sigma^2)$ |
| Densidad de probabilidad | $f_X(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$ para cualquier x real. |
| Probabilidad acumulada | No tiene expresión analítica |
| Parámetros | Parámetro de posición: $\mu \in (-\infty, \infty)$. Parámetro de escala: $\sigma > 0$. |
| Rango | $(-\infty, \infty)$ |
| Media | μ |
| Varianza | σ^2 |
| Modo | μ |
| Estimador | $\hat{\mu} = \bar{X}(n)$ $\hat{\sigma} = \sqrt{S^2(n)}$ |

| LOGNORMAL | $LN(\mu, \sigma^2)$ |
|-----------------------------------|---|
| Densidad de probabilidad | $f_X(x) = \begin{cases} \frac{1}{x \cdot \sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(\ln(x) - \mu)^2}{2 \cdot \sigma^2}} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ |
| Probabilidad acumulada | No tiene expresión analítica |
| Parámetros | Parámetro de escala: $\mu \in (-\infty, \infty)$. Parámetro de forma: $\sigma > 0$. |
| Rango | $[0, \infty)$ |
| Media | $e^{\mu + \frac{\sigma^2}{2}}$ |
| Varianza | $e^{2 \cdot \mu + \sigma^2} \cdot (e^{\sigma^2} - 1)$ |
| Modo | $e^{\mu - \sigma^2}$ |
| Estimador de máxima verosimilitud | $\hat{\mu} = \frac{\sum_{i=1}^n \ln(x_i)}{n}$ $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (\ln(x_i) - \hat{\mu})^2}{n}}$ |

| | |
|-------------|-------------------------------|
| BETA | beta (α_1, α_2) |
|-------------|-------------------------------|

| | |
|--------------------------|---|
| Densidad de probabilidad | $f_X(x) = \begin{cases} \frac{x^{\alpha_1-1} \cdot (1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{si } 0 < x < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$ <p>donde $B(\alpha_1, \alpha_2)$ es la función beta.</p> |
| Probabilidad acumulada | No tiene expresión analítica |
| Parámetros | Dos parámetros de forma: $\alpha_1 > 0, \alpha_2 > 0$. |
| Rango | $[0, 1)$ |
| Media | $\frac{\alpha_1}{\alpha_1 + \alpha_2}$ |
| Varianza | $\frac{\alpha_1 \cdot \alpha_2}{(\alpha_1 + \alpha_2)^2 \cdot (\alpha_1 + \alpha_2 + 1)}$ |
| Modo | $\frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$ si $\alpha_1 > 1$ y $\alpha_2 > 1$ 0 y 1 si $\alpha_1 < 1$ y $\alpha_2 < 1$ 0 si $(\alpha_1 < 1$ y $\alpha_2 \geq 1)$ o si $(\alpha_1 = 1$ y $\alpha_2 > 1)$ 1 si $(\alpha_1 \geq 1$ y $\alpha_2 < 1)$ o si $(\alpha_1 > 1$ y $\alpha_2 = 1)$ no está definido si $\alpha_1 = \alpha_2 = 1$ |
| Estimador | <p>Los parámetros de la distribución beta están relacionados con la media y la varianza de la distribución de la siguiente manera (Pedgen et al. 1995):</p> $\alpha_1 = \mu \cdot \left(\frac{(1-\mu) \cdot \mu}{\sigma^2} - 1 \right)$ $\alpha_2 = \mu \cdot (1 - \mu) \cdot \left((1 - \mu) \cdot \mu \cdot \sigma^2 - 1 \right)$ <p>Puede estimarse el valor de los parámetros sustituyendo la media y la varianza por sus estimadores: $\bar{X}(n)$ y $S^2(n)$ respectivamente.</p> |

| | |
|--------------------------|---|
| TRIANGULAR | $\text{triang}(a, b, c)$ |
| Densidad de probabilidad | $f_X(x) = \begin{cases} \frac{2 \cdot (x-a)}{(b-a) \cdot (c-a)} & \text{si } a \leq x \leq c \\ \frac{2 \cdot (b-x)}{(b-a) \cdot (b-c)} & \text{si } c < x \leq b \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{(x-a)^2}{(b-a) \cdot (c-a)} & \text{si } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a) \cdot (b-c)} & \text{si } c < x \leq b \\ 1 & \text{si } b < x \end{cases}$ |
| Parámetros | <p>a, b y c son números reales que satisfacen: $a < c < b$.</p> <p>Parámetro de posición, a.</p> <p>Parámetro de escala, $b - a$.</p> <p>Parámetro de forma, c.</p> |
| Rango | $[a, b]$ |
| Media | $\frac{a+b+c}{3}$ |
| Varianza | $\frac{a^2+b^2+c^2-a \cdot b-a \cdot c-b \cdot c}{18}$ |
| Modo | c |
| Estimadores | $\hat{a} = \min_{1 \leq i \leq n} \{x_i\}$ $\hat{b} = \max_{1 \leq i \leq n} \{x_i\}$ $\hat{c} = 3 \cdot \bar{X}(n) - \hat{a} - \hat{b}$ |

EQUIVALENCIAS ENTRE DISTRIBUCIONES

$$\begin{aligned}\text{beta}(1, 1) &= U(0, 1) \\ \text{beta}(1, 2) &= \text{triang}(0, 0, 1) \\ \text{beta}(2, 1) &= \text{triang}(0, 1, 1)\end{aligned}$$

$$\begin{aligned}\text{gamma}(1, \beta) &= \text{expo}(\beta) \\ \text{gamma}\left(\frac{k}{2}, 2\right) &= \chi_k^2 \\ \text{gamma}(m, \beta) &= \text{Erlang}(\beta, m) \quad \text{para } m \text{ entero positivo.}\end{aligned}$$

$$\text{Weibull}(1, \beta) = \text{expo}(\beta)$$

RELACIONES ENTRE DISTRIBUCIONES

$$X_i \sim \text{expo}(\beta) \quad \left| \quad Y = X_1 + \dots + X_m \quad \right| \rightarrow Y \sim \text{gamma}(m, \beta)$$

$$X_i \sim \text{gamma}(\alpha_i, \beta) \quad \left| \quad Y = X_1 + \dots + X_m \quad \right| \rightarrow Y \sim \text{gamma}(\alpha_1 + \dots + \alpha_m, \beta)$$

$$X_i \sim \text{gamma}(\alpha_i, \beta) \quad \left| \quad Y = \frac{X_1}{X_1 + X_2} \quad \right| \rightarrow Y \sim \text{beta}(\alpha_1, \alpha_2)$$

$$X_i \sim N(0, 1) \quad \left| \quad Y = X_1^2 + \dots + X_m^2 \quad \right| \rightarrow Y \sim \chi_m^2 \quad (\text{chi-cuadrado con } n \text{ grados de libertad})$$

$$X^\alpha \sim \text{expo}(\beta^\alpha) \leftrightarrow X \sim \text{Weibull}(\alpha, \beta)$$

$$X \sim \text{LN}(\mu, \sigma^2) \leftrightarrow \ln(X) \sim N(\mu, \sigma^2)$$

$$\begin{array}{l} X_1 \sim N(0, 1) \\ X_2 \sim \chi_m^2 \\ Y = \frac{X_1}{\sqrt{\frac{X_2}{m}}} \end{array} \quad \left| \quad \right. \rightarrow Y \sim t_m \quad (t \text{ de Student con } m \text{ grados de libertad}).$$

Apéndice C

Distribuciones de probabilidad discretas

| | |
|------------------|-------------------|
| BERNOULLI | Bernoulli (b) |
|------------------|-------------------|

| | |
|-----------------------------------|--|
| Aplicación | La variable aleatoria es el resultado de un experimento con dos posibles resultados: $\{0, 1\}$. La probabilidad de que resultado del experimento sea "1" es b . La probabilidad de que el resultado sea "0" es $1 - b$. |
| Probabilidad | $p_X(x) = \begin{cases} 1 - b & \text{si } x = 0 \\ b & \text{si } x = 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - b & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$ |
| Parámetros | $b \in (0, 1)$ |
| Rango | $\{0, 1\}$ |
| Media | b |
| Varianza | $b \cdot (1 - b)$ |
| Modo | $\begin{cases} 0 & \text{si } b < 0.5 \\ 0 \text{ y } 1 & \text{si } b = 0.5 \\ 1 & \text{si } b > 0.5 \end{cases}$ |
| Estimador de máxima verosimilitud | $\hat{b} = \bar{X}(n)$ |

| UNIFORME DISCRETA | $DU(i, j)$ |
|-----------------------------------|---|
| Aplicación | La variable aleatoria es el resultado de un experimento con $j-i+1$ posibles resultados equiprobables: $\{i, i+1, \dots, j\}$. |
| Probabilidad | $p_X(x) = \begin{cases} \frac{1}{j-i+1} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 0 & \text{si } x < i \\ \frac{\lfloor x \rfloor - i + 1}{j - i + 1} & \text{si } i \leq x \leq j \\ 1 & \text{si } x > j \end{cases}$ <p>donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x.</p> |
| Parámetros | <p>i, j son números enteros, tales que $i \leq j$.</p> <p>i: es un parámetro de posición.</p> <p>$j - i$: es un parámetro de escala.</p> |
| Rango | $\{i, i+1, \dots, j\}$ |
| Media | $\frac{i+j}{2}$ |
| Varianza | $\frac{(j-i+1)^2-1}{12}$ |
| Modo | No es único. |
| Estimador de máxima verosimilitud | $\hat{i} = \min \{x_1, \dots, x_n\}$ $\hat{j} = \max \{x_1, \dots, x_n\}$ |

| | |
|-----------------------------------|---|
| BINOMIAL | $\text{bin}(t, b)$ |
| Aplicación | Número de veces que se obtiene "1" como resultado de t experimentos independientes de Bernoulli. La probabilidad de obtener "1" en uno de los experimentos es b . |
| Probabilidad | $p_X(x) = \begin{cases} \binom{t}{x} \cdot b^x \cdot (1-b)^{t-x} & \text{si } x \in \{0, 1, \dots, t\} \\ 0 & \text{en cualquier otro caso} \end{cases}$ <p>donde el coeficiente binomial se define: $\binom{t}{x} = \frac{t!}{x!(t-x)!}$</p> |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{t}{i} \cdot b^i \cdot (1-b)^{t-i} & \text{si } 0 \leq x \leq t \\ 1 & \text{si } x > t \end{cases}$ <p>donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x.</p> |
| Parámetros | t es un número entero positivo. $b \in (0, 1)$ |
| Rango | $\{0, 1, \dots, t\}$ |
| Media | $t \cdot b$ |
| Varianza | $t \cdot b \cdot (1 - b)$ |
| Modo | $\begin{cases} b \cdot (t + 1) - 1 \text{ y } b \cdot (t + 1) & \text{si } b \cdot (t + 1) \text{ es un entero} \\ \lfloor b \cdot (t + 1) \rfloor & \text{en cualquier otro caso} \end{cases}$ |
| Estimador de máxima verosimilitud | Si t es conocido, entonces: $\hat{p} = \frac{\bar{X}(n)}{t}$ |

| | |
|-------------------|------------------|
| GEOMÉTRICA | $\text{geom}(b)$ |
|-------------------|------------------|

| | |
|-----------------------------------|---|
| Aplicación | Número de veces que sale "0" antes de que salga el primer "1" en una secuencia de experimentos de Bernoulli independientes, donde b es la probabilidad de que el resultado de un experimento sea "1". |
| Probabilidad | $p_X(x) = \begin{cases} b \cdot (1-b)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 1 - (1-b)^{\lfloor x \rfloor + 1} & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$ <p>donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x.</p> |
| Parámetros | $p \in (0, 1)$ |
| Rango | $\{0, 1, 2, \dots\}$ |
| Media | $\frac{1-b}{b}$ |
| Varianza | $\frac{1-b}{b^2}$ |
| Modo | 0 |
| Estimador de máxima verosimilitud | $\hat{p} = \frac{1}{X^{(n)}+1}$ |

| BINOMIAL NEGATIVA | $\text{negbin}(s, b)$ |
|-----------------------------------|---|
| Aplicación | Número de veces que sale "0" antes de obtener s veces el valor "1" en una secuencia de experimentos de Bernoulli independientes. La probabilidad de que el resultado de un experimento sea "1" es b . |
| Probabilidad | $p_X(x) = \begin{cases} \binom{s+x-1}{x} \cdot b^s \cdot (1-b)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} \sum_{i:0}^{\lfloor x \rfloor} \binom{s+i-1}{i} \cdot b^s \cdot (1-b)^s & \text{si } x \geq 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$ <p>donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x.</p> |
| Parámetros | s es un número entero positivo. $p \in (0, 1)$ |
| Rango | $\{0, 1, \dots\}$ |
| Media | $\frac{s \cdot (1-b)}{b}$ |
| Varianza | $\frac{s \cdot (1-b)}{b^2}$ |
| Modo | $\begin{cases} y \text{ e } y + 1 & \text{si } y \text{ es un entero} \\ \lfloor y \rfloor + 1 & \text{en cualquier otro caso} \end{cases}$ |
| Estimador de máxima verosimilitud | Si s es conocido: $\hat{p} = \frac{s}{X(n)+s}$. |

| POISSON | Poisson (λ) |
|-----------------------------------|---|
| Aplicación | Número de eventos que ocurren en un intervalo de tiempo, cuando el tiempo que transcurre entre eventos consecutivos están distribuido exponencialmente con frecuencia constante. |
| Probabilidad | $p_X(x) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^x}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en cualquier otro caso} \end{cases}$ |
| Probabilidad acumulada | $F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \cdot \sum_{i:0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{si } x \geq 0 \end{cases}$ <p>donde $\lfloor x \rfloor$ representa el mayor entero que es menor o igual que x.</p> |
| Parámetros | $\lambda > 0$ |
| Rango | $\{0, 1, 2, \dots\}$ |
| Media | λ |
| Varianza | λ |
| Modo | $\begin{cases} \lambda - 1 \text{ y } \lambda & \text{si } \lambda \text{ es un entero} \\ \lfloor \lambda \rfloor & \text{en cualquier otro caso} \end{cases}$ |
| Estimador de máxima verosimilitud | $\hat{\lambda} = \bar{X} (n)$ |

EQUIVALENCIAS ENTRE DISTRIBUCIONES

$$\begin{aligned} \text{Bernoulli}(b) &= \text{bin}(1, b) \\ \text{Bernoulli}(0.5) &= \text{DU}(0, 1) \end{aligned}$$

$$\text{geom}(b) = \text{negbin}(1, b)$$

RELACIONES ENTRE DISTRIBUCIONES

$$\begin{array}{l} X_i \sim \text{Bernoulli}(b) \\ Y = X_1 + \dots + X_t \end{array} \left| \rightarrow Y \sim \text{bin}(t, b) \right.$$

$$\begin{array}{l} X_i \sim \text{bin}(t_i, b) \\ Y = X_1 + \dots + X_m \end{array} \left| \rightarrow Y \sim \text{bin}(t_1 + t_2 + \dots + t_m, b) \right.$$

$$\begin{array}{l} X_i \sim \text{geom}(b) \\ Y = X_1 + \dots + X_s \end{array} \left| \rightarrow Y \sim \text{negbin}(s, b) \right.$$

$$\begin{array}{l} X_i \sim \text{negbin}(s_i, b) \\ Y = X_1 + \dots + X_m \end{array} \left| \rightarrow Y \sim \text{negbin}(s_1 + s_2 + \dots + s_m, b) \right.$$

$$\begin{array}{l} X_i \sim \text{Poisson}(\lambda_i) \\ Y = X_1 + \dots + X_m \end{array} \left| \rightarrow Y \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_m) \right.$$

Parte VI

Tablas

TABLA T.1
 Puntos críticos $t_{\nu, \gamma}$ para la distribución t , con ν grados de libertad. La última fila (donde $\nu = \infty$) son los puntos críticos, z_{γ} , de la distribución normal estándar.

| ν | 0.6000 | 0.7000 | 0.8000 | 0.9000 | 0.9333 | 0.9500 | 0.9600 | 0.9667 | 0.9750 | 0.9800 | 0.9833 | 0.9875 | 0.9900 | 0.9917 | 0.9938 | 0.9950 |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0.325 | 0.727 | 1.376 | 3.078 | 4.702 | 6.314 | 7.916 | 9.524 | 12.706 | 15.895 | 19.043 | 25.452 | 31.821 | 38.342 | 51.334 | 63.657 |
| 2 | 0.289 | 0.617 | 1.061 | 1.886 | 2.456 | 2.920 | 3.320 | 3.679 | 4.303 | 4.849 | 5.334 | 6.205 | 6.965 | 7.665 | 8.897 | 9.925 |
| 3 | 0.277 | 0.584 | 0.978 | 1.638 | 2.045 | 2.353 | 2.605 | 2.823 | 3.182 | 3.482 | 3.738 | 4.177 | 4.541 | 4.864 | 5.408 | 5.841 |
| 4 | 0.271 | 0.569 | 0.941 | 1.533 | 1.879 | 2.132 | 2.333 | 2.502 | 2.776 | 2.999 | 3.184 | 3.495 | 3.747 | 3.966 | 4.325 | 4.604 |
| 5 | 0.267 | 0.559 | 0.920 | 1.476 | 1.790 | 2.015 | 2.191 | 2.337 | 2.571 | 2.757 | 2.910 | 3.163 | 3.365 | 3.538 | 3.818 | 4.032 |
| 6 | 0.265 | 0.553 | 0.906 | 1.440 | 1.735 | 1.943 | 2.104 | 2.237 | 2.447 | 2.612 | 2.748 | 2.969 | 3.143 | 3.291 | 3.528 | 3.707 |
| 7 | 0.263 | 0.549 | 0.896 | 1.415 | 1.698 | 1.895 | 2.046 | 2.170 | 2.365 | 2.517 | 2.640 | 2.841 | 2.998 | 3.130 | 3.341 | 3.499 |
| 8 | 0.262 | 0.546 | 0.889 | 1.397 | 1.670 | 1.860 | 2.004 | 2.122 | 2.306 | 2.449 | 2.565 | 2.752 | 2.896 | 3.018 | 3.211 | 3.355 |
| 9 | 0.261 | 0.543 | 0.883 | 1.383 | 1.650 | 1.833 | 1.973 | 2.086 | 2.262 | 2.398 | 2.508 | 2.685 | 2.821 | 2.936 | 3.116 | 3.250 |
| 10 | 0.260 | 0.542 | 0.879 | 1.372 | 1.634 | 1.812 | 1.948 | 2.058 | 2.228 | 2.359 | 2.465 | 2.634 | 2.764 | 2.872 | 3.043 | 3.169 |
| 11 | 0.260 | 0.540 | 0.876 | 1.363 | 1.621 | 1.796 | 1.928 | 2.036 | 2.201 | 2.328 | 2.430 | 2.593 | 2.718 | 2.822 | 2.985 | 3.106 |
| 12 | 0.259 | 0.539 | 0.873 | 1.356 | 1.610 | 1.782 | 1.912 | 2.017 | 2.179 | 2.303 | 2.402 | 2.560 | 2.681 | 2.782 | 2.939 | 3.055 |
| 13 | 0.259 | 0.538 | 0.870 | 1.350 | 1.601 | 1.771 | 1.899 | 2.002 | 2.160 | 2.282 | 2.379 | 2.533 | 2.650 | 2.748 | 2.900 | 3.012 |
| 14 | 0.258 | 0.537 | 0.868 | 1.345 | 1.593 | 1.761 | 1.887 | 1.989 | 2.145 | 2.264 | 2.359 | 2.510 | 2.624 | 2.720 | 2.868 | 2.977 |
| 15 | 0.258 | 0.536 | 0.866 | 1.341 | 1.587 | 1.753 | 1.878 | 1.978 | 2.131 | 2.249 | 2.342 | 2.490 | 2.602 | 2.696 | 2.841 | 2.947 |
| 16 | 0.258 | 0.535 | 0.865 | 1.337 | 1.581 | 1.746 | 1.869 | 1.968 | 2.120 | 2.235 | 2.327 | 2.473 | 2.583 | 2.675 | 2.817 | 2.921 |
| 17 | 0.257 | 0.534 | 0.863 | 1.333 | 1.576 | 1.740 | 1.862 | 1.960 | 2.110 | 2.224 | 2.315 | 2.458 | 2.567 | 2.657 | 2.796 | 2.898 |
| 18 | 0.257 | 0.534 | 0.862 | 1.330 | 1.572 | 1.734 | 1.855 | 1.953 | 2.101 | 2.214 | 2.303 | 2.445 | 2.552 | 2.641 | 2.778 | 2.878 |
| 19 | 0.257 | 0.533 | 0.861 | 1.328 | 1.568 | 1.729 | 1.850 | 1.946 | 2.093 | 2.205 | 2.293 | 2.433 | 2.539 | 2.627 | 2.762 | 2.861 |
| 20 | 0.257 | 0.533 | 0.860 | 1.325 | 1.564 | 1.725 | 1.844 | 1.940 | 2.086 | 2.197 | 2.285 | 2.423 | 2.528 | 2.614 | 2.748 | 2.845 |
| 21 | 0.257 | 0.532 | 0.859 | 1.323 | 1.561 | 1.721 | 1.840 | 1.935 | 2.080 | 2.189 | 2.277 | 2.414 | 2.518 | 2.603 | 2.735 | 2.831 |
| 22 | 0.256 | 0.532 | 0.858 | 1.321 | 1.558 | 1.717 | 1.835 | 1.930 | 2.074 | 2.183 | 2.269 | 2.405 | 2.508 | 2.593 | 2.724 | 2.819 |
| 23 | 0.256 | 0.532 | 0.858 | 1.319 | 1.556 | 1.714 | 1.832 | 1.926 | 2.069 | 2.177 | 2.263 | 2.398 | 2.500 | 2.584 | 2.713 | 2.807 |
| 24 | 0.256 | 0.531 | 0.857 | 1.318 | 1.553 | 1.711 | 1.828 | 1.922 | 2.064 | 2.172 | 2.257 | 2.391 | 2.492 | 2.575 | 2.704 | 2.797 |
| 24 | 0.256 | 0.531 | 0.856 | 1.316 | 1.551 | 1.708 | 1.825 | 1.918 | 2.060 | 2.167 | 2.251 | 2.385 | 2.485 | 2.568 | 2.695 | 2.787 |
| 26 | 0.256 | 0.531 | 0.856 | 1.315 | 1.549 | 1.706 | 1.822 | 1.915 | 2.056 | 2.162 | 2.246 | 2.379 | 2.479 | 2.561 | 2.687 | 2.779 |
| 27 | 0.256 | 0.531 | 0.855 | 1.314 | 1.547 | 1.703 | 1.819 | 1.912 | 2.052 | 2.158 | 2.242 | 2.373 | 2.473 | 2.554 | 2.680 | 2.771 |
| 28 | 0.256 | 0.530 | 0.855 | 1.313 | 1.546 | 1.701 | 1.817 | 1.909 | 2.048 | 2.154 | 2.237 | 2.368 | 2.467 | 2.548 | 2.673 | 2.763 |
| 29 | 0.256 | 0.530 | 0.854 | 1.311 | 1.544 | 1.699 | 1.814 | 1.906 | 2.045 | 2.150 | 2.233 | 2.364 | 2.462 | 2.543 | 2.667 | 2.756 |
| 30 | 0.256 | 0.530 | 0.854 | 1.310 | 1.543 | 1.697 | 1.812 | 1.904 | 2.042 | 2.147 | 2.230 | 2.360 | 2.457 | 2.537 | 2.661 | 2.750 |
| 40 | 0.255 | 0.529 | 0.851 | 1.303 | 1.532 | 1.684 | 1.796 | 1.886 | 2.021 | 2.123 | 2.203 | 2.329 | 2.423 | 2.501 | 2.619 | 2.704 |
| 50 | 0.255 | 0.528 | 0.849 | 1.299 | 1.526 | 1.676 | 1.787 | 1.875 | 2.009 | 2.109 | 2.188 | 2.311 | 2.403 | 2.479 | 2.594 | 2.678 |
| 75 | 0.254 | 0.527 | 0.846 | 1.293 | 1.517 | 1.665 | 1.775 | 1.861 | 1.992 | 2.090 | 2.167 | 2.287 | 2.377 | 2.450 | 2.562 | 2.643 |
| 100 | 0.254 | 0.526 | 0.845 | 1.290 | 1.513 | 1.660 | 1.769 | 1.855 | 1.984 | 2.081 | 2.157 | 2.276 | 2.364 | 2.436 | 2.547 | 2.626 |
| ∞ | 0.253 | 0.524 | 0.842 | 1.282 | 1.501 | 1.645 | 1.751 | 1.834 | 1.960 | 2.054 | 2.127 | 2.241 | 2.326 | 2.395 | 2.501 | 2.576 |

TABLA T.2Puntos críticos, $\chi_{\nu,\gamma}^2$, de la distribución chi-cuadrado con ν grados de libertad.

| ν | γ | | | | | | |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 0.250 | 0.500 | 0.750 | 0.900 | 0.950 | 0.975 | 0.990 |
| 1 | 0.102 | 0.455 | 1.323 | 2.706 | 3.841 | 5.024 | 6.635 |
| 2 | 0.575 | 1.386 | 2.773 | 4.605 | 5.991 | 7.378 | 9.210 |
| 3 | 1.213 | 2.366 | 4.108 | 6.251 | 7.815 | 9.348 | 11.345 |
| 4 | 1.923 | 3.357 | 5.385 | 7.779 | 9.488 | 11.143 | 13.277 |
| 5 | 2.675 | 4.351 | 6.626 | 9.236 | 11.070 | 12.833 | 15.086 |
| 6 | 3.455 | 5.348 | 7.841 | 10.645 | 12.592 | 14.449 | 16.812 |
| 7 | 4.255 | 6.346 | 9.037 | 12.017 | 14.067 | 16.013 | 18.475 |
| 8 | 5.071 | 7.344 | 10.219 | 13.362 | 15.507 | 17.535 | 20.090 |
| 9 | 5.899 | 8.343 | 11.389 | 14.684 | 16.919 | 19.023 | 21.666 |
| 10 | 6.737 | 9.342 | 12.549 | 15.987 | 18.307 | 20.483 | 23.209 |
| 11 | 7.584 | 10.341 | 13.701 | 17.275 | 19.675 | 21.920 | 24.725 |
| 12 | 8.438 | 11.340 | 14.845 | 18.549 | 21.026 | 23.337 | 26.217 |
| 13 | 9.299 | 12.340 | 15.984 | 19.812 | 22.362 | 24.736 | 27.688 |
| 14 | 10.165 | 13.339 | 17.117 | 21.064 | 23.685 | 26.119 | 29.141 |
| 15 | 11.037 | 14.339 | 18.245 | 22.307 | 24.996 | 27.488 | 30.578 |
| 16 | 11.912 | 15.338 | 19.369 | 23.542 | 26.296 | 28.845 | 32.000 |
| 17 | 12.792 | 16.338 | 20.489 | 24.769 | 27.587 | 30.191 | 33.409 |
| 18 | 13.675 | 17.338 | 21.605 | 25.989 | 28.869 | 31.526 | 34.805 |
| 19 | 14.562 | 18.338 | 22.718 | 27.204 | 30.144 | 32.852 | 36.191 |
| 20 | 15.452 | 19.337 | 23.828 | 28.412 | 31.410 | 34.170 | 37.566 |
| 21 | 16.344 | 20.337 | 24.935 | 29.615 | 32.671 | 35.479 | 38.932 |
| 22 | 17.240 | 21.337 | 26.039 | 30.813 | 33.924 | 36.781 | 40.289 |
| 23 | 18.137 | 22.337 | 27.141 | 32.007 | 35.172 | 38.076 | 41.638 |
| 24 | 19.037 | 23.337 | 28.241 | 33.196 | 36.415 | 39.364 | 42.980 |
| 25 | 19.939 | 24.337 | 29.339 | 34.382 | 37.652 | 40.646 | 44.314 |
| 26 | 20.843 | 25.336 | 30.435 | 35.563 | 38.885 | 41.923 | 45.642 |
| 27 | 21.749 | 26.336 | 31.528 | 36.741 | 40.113 | 43.195 | 46.963 |
| 28 | 22.657 | 27.336 | 32.620 | 37.916 | 41.337 | 44.461 | 48.278 |
| 29 | 23.567 | 28.336 | 33.711 | 39.087 | 42.557 | 45.722 | 49.588 |
| 30 | 24.478 | 29.336 | 34.800 | 40.256 | 43.773 | 46.979 | 50.892 |
| 40 | 33.660 | 39.335 | 45.616 | 51.805 | 55.758 | 59.342 | 63.691 |
| 50 | 42.942 | 49.335 | 56.334 | 63.167 | 67.505 | 71.420 | 76.154 |
| 75 | 66.417 | 74.334 | 82.858 | 91.061 | 96.217 | 100.839 | 106.393 |
| 100 | 90.133 | 99.334 | 109.141 | 118.498 | 124.342 | 129.561 | 135.807 |

TABLA T.3

Valores críticos del test de Kolmogorov-Smirnov (Banks et al. 1996).

| Grados de libertad (<i>N</i>) | $D_{0.10}$ | $D_{0.05}$ | $D_{0.01}$ |
|------------------------------------|-------------------------|-------------------------|-------------------------|
| 1 | 0.950 | 0.975 | 0.995 |
| 2 | 0.776 | 0.842 | 0.929 |
| 3 | 0.642 | 0.708 | 0.828 |
| 4 | 0.564 | 0.624 | 0.733 |
| 5 | 0.510 | 0.565 | 0.669 |
| 6 | 0.470 | 0.521 | 0.618 |
| 7 | 0.438 | 0.486 | 0.577 |
| 8 | 0.411 | 0.457 | 0.543 |
| 9 | 0.388 | 0.432 | 0.514 |
| 10 | 0.368 | 0.410 | 0.490 |
| 11 | 0.352 | 0.391 | 0.468 |
| 12 | 0.338 | 0.375 | 0.450 |
| 13 | 0.325 | 0.361 | 0.433 |
| 14 | 0.314 | 0.349 | 0.418 |
| 15 | 0.304 | 0.338 | 0.404 |
| 16 | 0.295 | 0.328 | 0.392 |
| 17 | 0.286 | 0.318 | 0.381 |
| 18 | 0.278 | 0.309 | 0.371 |
| 19 | 0.272 | 0.301 | 0.363 |
| 20 | 0.264 | 0.294 | 0.356 |
| 25 | 0.24 | 0.27 | 0.32 |
| 30 | 0.22 | 0.24 | 0.29 |
| 35 | 0.21 | 0.23 | 0.27 |
| >35 | $\frac{1.22}{\sqrt{N}}$ | $\frac{1.36}{\sqrt{N}}$ | $\frac{1.63}{\sqrt{N}}$ |

Bibliografía

- Banks, J., Carson, J. S. & Nelson, B. L. (1996), *Discrete-Event System Simulation*, Prentice-Hall.
- Box, G. E. P., Hunter, W. G. & Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, John Wiley & Sons.
- Bratley, P., Fox, B. L. & Schrage, L. E. (1987), *A Guide to Simulation*, Springer.
- Buss, A. (2001), 'Basic event graph modeling', *Simulation News Europe* **31**(April), 1–6.
- Cassandras, C. G. & Lafortune, S. (1999), *Introduction to Discrete Event Systems*, Kluwer Academic Publishers.
- Cellier, F. C. (1991), *Continuous System Modeling*, Springer-Verlag.
- Fishman, G. S. (1978), *Principles of Discrete Event Simulation*, John Wiley & Sons.
- Fishman, G. S. (1995), *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag.
- Franta, W. R. (1977), *The Process View of Simulation*, North Holland, New York.
- Guasch, A., Piera, M. A., Casanovas, J. & Figueras, J. (2002), *Modelado y Simulación. Aplicación a Procesos Logísticos de Fabricación y Servicios*, Editions UPC.
- Hicks, C. R. & Turner, K. V. (1999), *Fundamental Concepts in the Design of Experiments*, Oxford University Press.
- Hoover, S. V. & Perry, R. F. (1989), *Simulation. A Problem-Solving Approach*, Addison-Wesley Publishing.
- HotBits*, (Fourmilab, Suiza) (2003). <http://www.fourmilab.ch/hotbits/>.
- Kelton, W. D., Sadowski, R. P. & Sadowski, D. A. (2002), *Simulation with Arena*, McGraw-Hill.
- Kobayashi, H. (1978), *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology*, Addison-Wesley Publishing Company.
- Law, A. M. & Kelton, W. D. (2000), *Simulation Modeling and Analysis*, McGraw-Hill.
- Ljung, L. & Torkel, G. (1994), *Modeling of Dynamic Systems*, Prentice-Hall.
- Mersenne Twister*, (Universidad de Keio, Japón) (2003). <http://www.math.keio.ac.jp/matumoto/emt.html>.
- Montgomery, D. C. (2001), *Design and Analysis of Experiments*, John Wiley & Sons.
- Neelamkavil, F. (1987), *Computer Simulation and Modelling*, John Wiley & Sons.
- Papoulis, A. (1991), *Probability, Random Variables and Stochastic Processes*, McGraw-Hill.
- Pedgen, C. D., Shannon, R. E. & Sadowsky, R. P. (1995), *Introduction to Simulation Using SIMAN*, McGraw-Hill.

- Pidd, M. (1989), *Computer Modelling for Discrete Simulation*, John Wiley & Sons.
- Press, W. H. (1992), *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press.
- Proyecto pLab, (Universidad de Salzburgo)* (2003). <http://random.mat.sbg.ac.at/>.
- RAND Corporation* (2003). <http://www.rand.org/publications/classics/randomdigits/>.
- Random.org, (Universidad de Dublín, Irlanda)* (2003). <http://www.random.org/>.
- Rockwell (2000a), *Arena Standard Edition. User's Guide*, Rockwell Software Inc.
- Rockwell (2000b), *Arena Standard Edition. Variables Guide*, Rockwell Software Inc.
- Ross, S. M. (1990), *A Course in Simulation*, Macmillan Publishing Company.
- Ross, S. M. (1994), *A First Course in Probability*, Prentice-Hall.
- Rubinstein, R. Y. (1981), *Simulation and the Monte Carlo Method*, John Wiley & Sons.

Índice alfabético

- abstracción, 11
- aceptación/rechazo, 192
- acumuladores estadísticos, 34, 36
- Arena, 15
- atributo, 33
- autocorrelación, 180
 - no autocorrelación, 181
 - positiva, 181
- balking, 91
- Bonferroni, desigualdad de, 220
- Bonferroti, desigualdad de, 214
- box & whiskers, 283
- boxplot, 283
- calendario de eventos, 43
- carrera, 175
 - creciente, 176
 - decreciente, 176
 - longitud, 175
- coeficiente de correlación, 272
- coeficiente de variación, 126
- cola, 34
- comportamiento
 - estacionario, 205
 - transitorio, 205
- composición, 191
- convolución, 192
- corrección del sesgo, 158
- correlación
 - negativa, 235, 272
 - positiva, 235, 272
- covarianza, 235, 271
 - estacionaria, 273
- cuantil, 129, 269
- datos
 - de entrada, 14
- densidad de probabilidad, 269
- desviación estándar, 271
- disciplina de la cola, 109
- discretización temporal, 7
- distribución
 - binomial, 125
 - de Bernoulli, 125
 - de Poisson, 126
 - de Weibull, 125
 - en el estacionario, 205
 - exponencial, 125
 - gamma, 125
 - geométrica, 126
 - negativa binomial, 126
 - normal, 125
 - triangular, 125
 - uniforme, 125
 - uniforme discreta, 125
- división por rebosamiento, 163
- division simulada, 165
- efecto cruzado, 250
- efecto principal, 249
- entidad, 33, 81
 - actividad, 34
 - instanciación, 33
 - realización, 33
- entornos de modelado, 15
- espacio muestral, 268
- evento, 34
 - calendario, 35
 - en el estado, 35
 - en el tiempo, 35
- eventos simultáneos, 65
- experimento, 4
 - condición final, 37
 - condición inicial, 37
 - de Bernoulli, 125, 278
 - de screening, 245, 254
 - definición, 32
 - diseño, 14
 - diseño de, 243
 - factor, 243
 - factor cualitativo, 247
 - factor cuantitativo, 247
 - factorial completo, 248
 - interacción, 244
 - nivel de un factor, 244
 - optimización, 243
 - respuesta, 243
 - respuesta categórica, 246
 - respuesta continua, 246
 - respuesta discreta, 246
- expresión, 93
- factor
 - manipulable, 248
 - no manipulable, 248

- GCL, 161
 - mixto, 163
 - mixtos, 162
 - multiplicativo, 164
 - multiplicativo de módulo primo, 165
 - multiplicativos, 162
- generación observaciones
 - distribución binomial, 195
 - distribución de Bernoulli, 195
 - distribución de Poisson, 195
 - distribución discreta empírica, 196
 - distribución discreta uniforme, 195
 - distribución empírica, 194
 - distribución exponencial, 193
 - distribución geométrica, 195
 - distribución lognormal, 193
 - distribución negativa binomial, 195
 - distribución normal, 193
 - distribución triangular, 193
 - distribución uniforme, 193
 - proceso de llegada, 196
 - proceso de Poisson con llegadas en grupo, 197
 - proceso de Poisson estacionario, 196
 - proceso de Poisson no estacionario, 197
- generador combinado, 166
- generador congruencial cuadrático, 166
- generador de Fibonacci, 166
- generador de Tausworthe, 168
- generador recursivo combinado múltiple, 168
- gráfica Q-Q, 129
- grafo de eventos, 63
 - arco dirigido, 63
 - arco incondicional, 64
 - condición del arco, 63
 - evento Run, 64
 - nodo, 63
- hipótesis nula, 280
- hipótesis paramétrica, 280
- implementación, 17
- independencia estadística, 36
- intensidad del tráfico, 109
- interacción de tercer orden, 253
- interacción entre factores, 250
- lenguajes
 - de modelado, 15
 - de programación, 15
 - de simulación, 15
- método experimental, 4
- marco experimental, 8
- Markov
 - cadena, 104
 - propiedad de, 105
 - sistema de, 104
- media, 270
- media en el estacionario
 - método de la media por lotes, 212, 213
 - método de réplicas/eliminación, 212, 213
- mediana, 269
- modelado
 - agregación de recursos, 13
 - agregación temporal, 13
 - aproximaciones, 13
 - jerárquico, 12
 - metodología, 11
 - modular, 12
 - omisión, 13
 - orientado a eventos, 15, 38, 41
 - orientado a procesos, 15, 38, 40, 41, 81
 - sustitución, 13
- modelo, 3
 - aspectos matemáticos, 35
 - componentes, 32
 - de tiempo continuo, 7
 - de tiempo discreto, 7
 - determinista, 6
 - diseño, 12
 - documentación, 17
 - estocástico, 6, 14
 - estructura lógica, 32
 - físico, 5
 - híbridos, 7
 - matemático, 5, 6
 - mental, 5
 - parametrizar, 33
 - reutilización, 17
 - traducción, 15
 - validación, 15, 16
 - verbal, 5
 - verificación, 15, 16
- número aleatorio, 153
- números pseudo aleatorios, 159
 - independencia, 160
 - reproducibilidad, 161
 - secuencia, 161
 - uniformidad, 160
- nivel de significación, 171
- OptQuest, 257
 - control, 258
 - respuesta, 258
 - restricción, 258
- Output Analyzer, 223
 - instalación, 223
- p-value, 149
- parámetro
 - de escala, 120
 - de forma, 120

- de posición, 120
- percentil, 269
- periodo de un generador, 162
 - periodo completo, 163
- proceso de llegada, 109
- Process Analyzer (PAN), 225
 - escenario, 225
 - proyecto, 225
- recurso, 34
 - individual, 34
 - unidad de, 34
- reducción de la varianza, 234
 - números pseudo aleatorios comunes, 234
 - variables aleatorias antitéticas, 240
- región experimental, 248
- reloj de la simulación, 33, 35, 43
- rutina
 - de evento, 43
 - de informes, 43
 - de tiempo, 43
 - inicialización, 43
- sesgo, 127
- SIMAN, 15
- simplificación, 11
- simulación, 6
 - con condición de finalización, 205, 208
 - con terminación, 17
 - de Monte Carlo, 21
 - en el estacionario, 209
 - sin condición de finalización, 205, 209
 - sin terminación, 17
- sistema, 3
- Sturges, regla de, 128
- submodelo, 94
- superficie de respuesta, 256
- supremo, 140
- test
 - característica de operación, 281
 - carreras, 175
 - chi-cuadrado, 136, 174
 - Diehard, 171
 - empírico, 170
 - frecuencia, 172
 - independencia, 170, 175
 - Kolmogorov-Smirnov, 136, 173
 - Kruskal-Wallis, 124
 - nivel de significación, 281
 - potencia, 281
 - serie, 172
 - uniformidad, 170, 172
- tiempo de servicio, 109
- transformación inversa, 188
- utilización, 89
- utilización planificada, 89
- valor esperado, 270
- variable, 33, 93
 - de estado, 37
 - de salida, 36
- variable aleatoria, 268
 - continua, 268
 - discreta, 268
 - generación de observaciones, 268
- Welch, metodo de, 209