



FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA E
INVESTIGACIÓN OPERATIVA

**MODELOS Y MÉTODOS DE SIMULACIÓN:
Técnicas de Monte Carlo.
APLICACIONES EN FINANZAS**

Begoña Vitoriano

bvitoriano@mat.ucm.es

Universidad Complutense de Madrid

ÍNDICE

I MODELADO DE SISTEMAS MEDIANTE SIMULACIÓN	5
I.1 SISTEMAS, MODELOS Y SIMULACIÓN	5
I.2 VENTAJAS E INCONVENIENTES DE LA SIMULACIÓN	8
I.3 TIPOS DE SISTEMAS Y DE SIMULACIÓN	9
I.4 ELEMENTOS DE UN MODELO DE SIMULACIÓN DINÁMICO	11
<i>I.4.1 El reloj de simulación.....</i>	<i>11</i>
<i>I.4.2 Mecanismo de transición.....</i>	<i>12</i>
I.5 ORGANIZACIÓN DE UN MODELO DE SIMULACIÓN DINÁMICO	13
I.6 ALGUNAS APLICACIONES DE SIMULACIÓN	18
<i>I.6.1 Disponibilidad dinámica de un sistema paralelo</i>	<i>18</i>
<i>I.6.2 Fiabilidad estática de un sistema de generación de energía eléctrica.....</i>	<i>19</i>
<i>I.6.3 Fiabilidad dinámica de un sistema de energía eléctrica.....</i>	<i>21</i>
I.7 ETAPAS EN EL DESARROLLO DE UN ESTUDIO DE SIMULACIÓN	23
II MODELADO DE LA ALEATORIEDAD EN SIMULACIÓN	27
II.1 REVISIÓN DE PROBABILIDAD Y ESTADÍSTICA	27
<i>II.1.1 Variables aleatorias y sus propiedades</i>	<i>27</i>
<i>II.1.2 Procesos estocásticos.....</i>	<i>33</i>
<i>II.1.3 Estimación puntual.....</i>	<i>35</i>
<i>II.1.4 Estimación por intervalos de confianza.....</i>	<i>37</i>
<i>II.1.5 Contrastes o tests de hipótesis</i>	<i>38</i>
II.2 IDENTIFICACIÓN DE PATRONES.....	40
<i>II.2.1 Distribuciones empíricas</i>	<i>41</i>
<i>II.2.2 Distribuciones teóricas</i>	<i>43</i>
II.3 GENERACIÓN DE VARIABLES ALEATORIAS	51
<i>II.3.1 Generación de muestras uniformes.....</i>	<i>51</i>

II.3.2	Generación de variables aleatorias discretas.....	56
II.3.2.1	Método general o estándar.....	56
II.3.2.2	Binomial (n, p)	58
II.3.2.3	Geométrica (p)	58
II.3.3	Generación de variables aleatorias absolutamente continuas.....	59
II.3.3.1	Método de la transformada inversa.....	59
II.3.3.2	Método de aceptación - rechazo.....	60
II.3.3.3	Aplicación a una discreta: Poisson(λ)	63
II.3.3.4	Normal (μ, σ)	64
II.3.3.5	Lognormal (μ, σ)	66
II.3.3.6	$\gamma(p, a)$ y Beta (α, β)	66
II.3.3.7	χ_n^2	68
II.3.3.8	t_n de Student	68
II.3.3.9	$F_{m,n}$ de Fisher-Snedecor.....	69
II.3.4	Generación de variables aleatorias mixtas	69
II.3.5	Generación de variables aleatorias multidimensionales	70
II.3.5.1	Nociones de cálculo matricial	70
II.3.5.2	Normal Multivariante $N(\mu, \Sigma)$	72
II.3.5.3	Multinomial	74
II.3.5.4	T_n de Student multivariante	75
III	SOFTWARE DE SIMULACIÓN DE MODELOS DINÁMICOS	77
III.1	LENGUAJES DE SIMULACIÓN VERSUS LENGUAJES DE PROPÓSITO GENERAL	77
III.2	TIPOS DE SOFTWARE DE SIMULACIÓN Y CARACTERÍSTICAS	78
III.3	ENFOQUES O ESTRATEGIAS.....	80
IV	ANÁLISIS DE RESULTADOS	85
IV.1	COMPORTAMIENTOS TRANSITORIO Y ESTACIONARIO DE UN PROCESO ESTOCÁSTICO ...	85

IV.2 TIPOS DE SIMULACIÓN SEGÚN EL ANÁLISIS DE RESULTADOS	87
IV.3 ESTIMACIÓN DE VARIABLES RESPUESTA: PRECISIÓN Y TAMAÑO MUESTRAL	88
IV.4 ESTIMACIÓN DE PARÁMETROS ESTACIONARIOS: EL PROBLEMA DEL ESTADO INICIAL TRANSITORIO.....	90
IV.5 ESTIMACIÓN DE INTEGRALES POR MONTE CARLO	93
IV.5.1 Método Monte Carlo de acertar o fallar.....	93
IV.5.2 Método de Monte Carlo crudo.....	94
IV.6 TÉCNICAS DE REDUCCIÓN DE LA VARIANZA	94
IV.6.1 Muestreo correlado (números aleatorios comunes)	95
IV.6.2 Variables antitéticas	96
IV.6.3 Variables de control.....	97
IV.6.4 Condicionamiento	98
IV.6.5 Muestreo estratificado	99
IV.6.6 Muestreo por importancia	100
IV.7 DISEÑO DE EXPERIMENTOS	101
IV.7.1 Principios del diseño experimental.....	101
IV.7.2 Diseño con un factor.....	104
IV.7.3 Diseño en bloques aleatorizados	106
IV.7.4 Modelos factoriales: con dos factores y con interacción.....	108
IV.7.5 Modelos factoriales con más de dos factores	110
IV.7.6 Cuadrados latinos.....	110
IV.7.7 Modelos con efectos aleatorios: Diseños jerárquicos	114
IV.7.8 Diseños factoriales a dos niveles (2^k).....	116
IV.7.9 Modelo de regresión	124
V CONSTRUCCIÓN DE MODELOS DE SIMULACIÓN VÁLIDOS Y CREÍBLES .	127
V.1 DEFINICIONES.....	127
V.2 METODOLOGÍA DE VALIDACIÓN Y CREDIBILIDAD	129
V.2.1 Comparación entre datos reales y simulados	130

VI APLICACIONES DE LA SIMULACIÓN EN FINANZAS	133
VI.1 NOCIONES BÁSICAS SOBRE DERIVADOS FINANCIEROS.....	133
VI.1.1 <i>Contratos de futuros</i>	134
VI.1.2 <i>Opciones</i>	135
VI.1.2.1 Opciones “vainilla” (<i>vanilla options</i>).....	135
VI.1.2.2 Opciones “exóticas” (<i>exotic options</i>).....	138
VI.2 VALORACIÓN DE OPCIONES Y FUTUROS	140
VI.2.1 <i>Conceptos básicos en valoración de derivados</i>	140
VI.2.2 <i>Estimación del valor de opciones sin posibilidad de ejercicio anticipado mediante simulación</i>	146
VI.2.3 <i>Estimación del valor de opciones con posibilidad de ejercicio anticipado mediante métodos numéricos</i>	149
VI.2.3.1 Método binomial.....	152
VII REFERENCIAS.....	157
VIII BIBLIOTECA DE PROBLEMAS.....	159
VIII.1 PROBLEMAS DE MÉTODOS DE MONTE CARLO	159
VIII.2 RESULTADOS DE LOS PROBLEMAS DE MÉTODOS DE MONTE CARLO	163
VIII.3 PROBLEMAS DE SIMULACIÓN DE SISTEMAS DINÁMICOS.....	169
VIII.4 RESULTADOS DE LOS PROBLEMAS DE SIMULACIÓN DE SISTEMAS DINÁMICOS	181
IX PRÁCTICAS PROPUESTAS PARA FINANZAS.....	209

I Modelado de sistemas mediante simulación

I.1 Sistemas, modelos y simulación

La primera vez en la historia que se habló de simulación fue en el año 1949 cuando John Von Neumann y Stanislaw Ulam presentaron el denominado método de Monte Carlo. Desde entonces la simulación ha sufrido un crecimiento muy fuerte y, especialmente en las dos últimas décadas, este crecimiento ha sido vertiginoso gracias al desarrollo de los ordenadores.

Dar una definición exacta de la simulación no es una tarea fácil dada la amplitud de las aplicaciones y sistemas a los que se aplica. Sin embargo, una buena definición sería la dada por Shannon en 1975, según la cual, simulación es el proceso de diseñar un modelo de un sistema real y llevar a cabo experiencias con él, con la finalidad de aprender el comportamiento del sistema o de evaluar diversas estrategias para el funcionamiento del sistema.

Varios conceptos son utilizados en esta definición y deben ser precisados.

- *Sistema*: Conjunto de objetos o ideas que están interrelacionadas entre sí como una unidad para la consecución de un fin. Forma parte de la vida real.
- *Modelo*: Representación simplificada de un sistema. Es una abstracción del sistema.

El objetivo es crear un modelo del sistema a partir de la observación. Sin embargo, esto no es particular de la simulación ya que en todo análisis de sistemas éste es un paso necesario, lo que es particular es el tipo de modelo y el uso que se hace de él.

Así, se dice que existen dos procedimientos para obtener modelos de sistemas:

- *Análisis teórico o método deductivo*: con este procedimiento se lleva a cabo un estudio cualitativo de los fenómenos que caracterizan el comportamiento del sistema, que son plasmados en relaciones matemáticas concretas que definen las ecuaciones descriptivas del proceso. Para dar una respuesta con este tipo de modelos hay que resolver las ecuaciones descriptivas del proceso, tarea que en ocasiones es especialmente difícil.

Como ejemplo, supóngase un objeto que se mueve con una cierta trayectoria, una velocidad en cada momento y una aceleración, que a su vez dependen de otras condiciones externas como las características del terreno en que se encuentre o la cercanía de otros objetos similares. Para dar respuesta, por ejemplo al punto en que se encontrará pasado un determinado tiempo con este tipo de modelo, hay que resolver las ecuaciones del movimiento del objeto, lo cual no es inmediato si el objeto cambia su trayectoria y

aceleración con el tiempo, las características de la posición, el movimiento de otros objetos, etc.

- *Análisis experimental o método inductivo*: consiste en construir un modelo matemático a partir de medidas realizadas sobre el sistema, dando una descripción detallada de cómo evoluciona a lo largo del tiempo, con el fin de observar el comportamiento del modelo y llevar a cabo experiencias con él: simulación del modelo.

Con el mismo ejemplo del objeto móvil, se puede ir representando en pequeños intervalos de tiempo las variaciones de velocidad, aceleración, posición, mediante la definición directa que relaciona a estas variables, tanto del móvil observado como de los cercanos, modificando en cada momento los parámetros según la posición en que esté él y los demás objetos, hasta llegar a cumplir el tiempo que se desea observar. Con este modelo es posible ensayar diversas respuestas ante distintas situaciones.

Así pues, con este procedimiento el objetivo no es conocer el sistema en sí, sino su comportamiento ante diversas situaciones ¡Los modelos de simulación no se resuelven, se hacen funcionar!

Ejemplo 1: Se desea construir una carretera y para ello se ha de hacer un túnel a través de una montaña. Existen dos puntos posibles donde hacer el túnel, correspondientes a dos montañas distintas M1 y M2 cercanas. Supongamos que se tiene que decidir si hacerlo en un punto o en otro. Mediante estudios preliminares se sabe que en el punto M1 la longitud del túnel habría de ser L_1 y en la montaña M2, L_2 . En la primera de ellas, por las características del terreno, se perforaría a razón de x_1 unidades por jornada de trabajo, mientras que en la otra montaña sería a razón de x_2 unidades.

La empresa debe recibir una maquinaria nueva con una probabilidad 0.71. La probabilidad de que la nueva maquinaria se averíe en M1 es 0.14 y en M2 es 0.16. Para la maquinaria vieja estas probabilidades son, respectivamente, 0.28 y 0.19.

Las averías pueden ser de dos tipos: graves con probabilidad 0.35 y 4 jornadas de trabajo de reparación o leves con 1 jornada de trabajo de reparación.

La pregunta es: ¿dónde se debe perforar el túnel para tardar lo menos posible en la construcción de la carretera?

Un modelo de simulación desarrollado para la perforación en el punto M1 sería, por ejemplo, el de la figura 1.

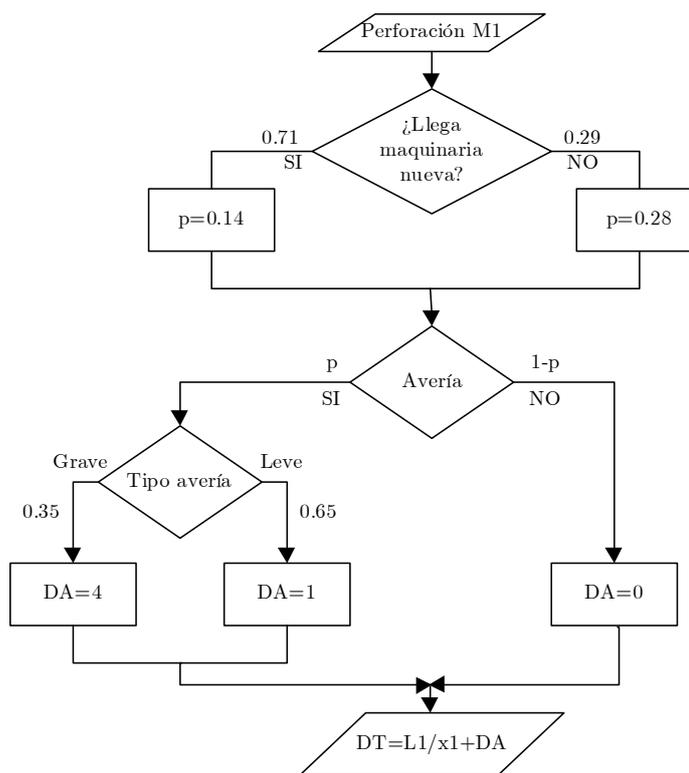


Figura 1. Modelo de simulación de perforación montaña M1.

Análogamente, se haría otro para el punto M2. Este modelo se programa en un lenguaje general y luego es simulado en el ordenador varias veces de modo que se obtenga una muestra simulada de la duración de la perforación en cada uno de los puntos. Una vez obtenida la muestra la decisión puede ser tomada eligiendo el punto que tenga un menor valor de la duración esperada. Sin embargo, este cálculo también puede ser obtenido mediante el método teórico, es decir, mediante cálculo de probabilidades. A modo de ejemplo, con el método teórico se obtuvo una duración esperada en M1 de 19.37023 jornadas y en M2 de 20.345835 y después de 50 simulaciones se obtuvo una duración media en M1 de 19.34 jornadas y en M2 de 20.22 jornadas.

De este ejemplo, también se deduce la necesidad de un mecanismo eficaz para generar variables aleatorias que sea rápido, dado que hay que repetir la simulación un número elevado de veces para dar por válidos los resultados obtenidos.

Ejemplo 2: A Pepe, su mejor amigo Juan le propone un juego: lanzar un dado hasta que la diferencia entre las veces que han salido números pares e impares sea tres. Pepe pagará 100

monedas en cada tirada y cobrará 1000 al final. ¿Cómo puede comprobar Pepe si Juan es realmente un buen amigo?

Para comprobarlo, Pepe puede hacer los cálculos necesarios para calcular la esperanza de la ganancia o de la pérdida o puede simular las tiradas del dado y observar los resultados.

I.2 Ventajas e inconvenientes de la simulación

Puesto que existen dos métodos para obtener modelos, cabe preguntarse cuándo es ventajoso utilizar la simulación y qué inconvenientes puede tener.

Cuándo puede ser ventajosa la simulación:

- Si no existe formulación matemática del modelo o los métodos de resolución. Por ejemplo, cuando se va a construir un aeropuerto para prever las necesidades de infraestructuras necesarias no existe un modelo matemático que pueda contemplar todo conjuntamente. Otro ejemplo son los sistemas de líneas de espera para los que se puede plantear un modelo, pero para muchos de ellos no existen métodos matemáticos para resolver las ecuaciones resultantes.
- Sí existen, pero resulta más barato y sencillo simular, ya que en muchas ocasiones el modelo y su resolución resultan más sencillos.
- Si se desea experimentar con el sistema antes de su uso o construcción. El ejemplo más conocido son los simuladores de vuelo, pero no es el único, ya que cada vez es más habitual la implantación de puestos de formación para operadores de sistemas, de modo que puedan practicar con el modelo como si lo hicieran con el sistema.
- Si es imposible experimentar sobre el sistema y se desean prevenir eventualidades. Éste es uno de los usos más habituales, la prevención de eventualidades, de modo que si posteriormente se da alguna de ellas se tenga identificada la respuesta óptima ante la eventualidad. Uno de los casos más representativos es el de la simulación de vuelos espaciales, de modo que antes de lanzarlo se ha simulado el sistema para tener respuestas en caso de fallo de algún mecanismo o similar. Obviamente, el éxito de este uso pasa por una buena identificación de las posibles eventualidades.
- Si hay razones éticas que impiden la experimentación, como en el caso de sistemas biológicos humanos.

- Si se desea reducir escalas de tiempo, pues la evolución del sistema es muy lenta. Por ejemplo en el estudio de diferentes políticas de talas de árboles no es viable esperar a ver cuál es la evolución del bosque con una determinada política, no sólo por las consecuencias que pueda tener, sino porque para observarlas habría que esperar numerosos años.
- Por último, una característica importante de la simulación es que permite estudiar sistemas dinámicos en tiempo real.

Sin embargo, también tiene ciertos inconvenientes la simulación:

- La construcción del modelo puede ser compleja y costosa. Por ejemplo, la construcción de un buen modelo socioeconómico mundial puede llevar unos cinco años de trabajo a un equipo.
- Es frecuente despreciar elementos o relaciones sin importancia aparente y obtener resultados falsos, aunque este peligro existe en cualquier proceso de desarrollo de un modelo, no sólo en los modelos de simulación.
- Es difícil establecer el grado de precisión de los resultados, ya que se obtienen muestras y como tales han de ser analizadas, con sus limitaciones. Es decir, cuando existe aleatoriedad los resultados han de verse como tales, aleatorios, y analizados con sumo cuidado y rigurosidad mediante técnicas estadísticas.

I.3 Tipos de sistemas y de simulación

A la hora de obtener un modelo de simulación de un sistema dinámico es fundamental definir el estado del sistema. Se llama estado de un sistema al conjunto de variables necesarias para describir un sistema en un instante particular de tiempo relativo a los objetivos de un estudio.

Los sistemas se clasifican a partir de sus variables de estado:

- *Sistemas continuos*: Las variables de estado cambian de forma continua con el tiempo. Por ejemplo, el movimiento de un tren por una vía, las variables de estado son posición, velocidad y aceleración.
- *Sistemas discretos*: Las variables de estado cambian en ciertos instantes de tiempo. Por ejemplo, un sistema de atención a clientes atendido por un solo servidor, la variable de estado es el número de clientes que hay en el centro de servicio.

Los modelos de simulación a su vez pueden ser clasificados atendiendo a varios criterios:

- Según la *evolución del tiempo*
 - *Estáticos*: representan un sistema en un instante particular. A menudo, a este tipo de simulación se la denomina simulación de Monte Carlo.
 - *Dinámicos*: representan un sistema que evoluciona con el tiempo.
 - Según la *aleatoriedad*
 - *Deterministas*: no incluyen variables aleatorias. Dados unos datos de entrada, existe un único conjunto posible de datos de salida.
 - *Probabilistas* o *estocásticos*: contienen variables aleatorias, las salidas son aleatorias (estimaciones de las verdaderas características).
 - Según las *variables de estado*
 - *Continuos*: si todas las variables de estado cambian de forma continua con el tiempo.
 - *Discretos*: si todas las variables de estado cambian en determinados instantes de tiempo. Se definen como *eventos* aquellos sucesos que pueden producir un cambio en el estado del sistema. A estos modelos también se les llama *modelos de simulación de eventos discretos*.
- Hay que tener en cuenta, sin embargo, que no siempre de un sistema discreto se hace un modelo discreto, hay casos en que por simplificar es mejor tratarlo como un sistema continuo. Un caso muy habitual son los modelos de tráfico, en que aunque los cambios son discretos, modelarlos como flujos continuos da mucho mejor resultado.
- *Híbridos* o *combinados*: si incluyen variables de estado continuas y discretas.

Un ejemplo de un modelo continuo es el clásico modelo presa - depredador de Lotka–Volterra. En él se supone una población (grande) y cerrada (no hay migraciones) de presas y depredadores. El objetivo es estudiar cómo varía el número de cada uno de ellos en el tiempo. El modelo propuesto por Lotka–Volterra sería el siguiente:

$X(t)$: número individuos presa en el instante t

$Y(t)$: número de individuos depredador en el instante t

$$\frac{dX}{dt} = rX(t) - aX(t)Y(t), \quad a > 0$$
$$\frac{dY}{dt} = -sY(t) + bX(t)Y(t), \quad b > 0$$

donde r es la tasa de crecimiento de la población presa en ausencia de depredadores, s la tasa de decrecimiento de los individuos depredador en ausencia de presas, y a y b son parámetros para ponderar la presencia de los otros individuos en la variación de población. Este modelo se puede intentar resolver mediante métodos clásicos, sin embargo, es mucho más sencillo simularlo tomando pequeños intervalos de tiempo y actualizando las variables de estado en esos intervalos mediante las ecuaciones descritas (directamente al ser pequeños intervalos). Obsérvese que desde un punto de vista teórico el sistema presentado sería discreto. Sin embargo, cuando las poblaciones son suficientemente grandes, el modelo apropiado es un modelo continuo.

En este documento nos vamos a centrar en los modelos de simulación *dinámicos*, *estocásticos* (el caso determinista es considerado un caso particular) y *discretos*: *modelos de simulación de eventos discretos*.

I.4 Elementos de un modelo de simulación dinámico

I.4.1 El reloj de simulación

En un modelo de simulación dinámico el elemento característico es el tiempo y su avance. Para ello se utiliza una variable que registra la cantidad de tiempo que ha sido simulada: el *reloj de simulación*. Esta variable no representa tiempo real, no es tiempo de ejecución, sino que es un contador interno del modelo.

El reloj de simulación hay que “moverlo”, incrementar su valor. Para ello hay dos métodos:

- *Incremento en tiempo fijo (time step)*: el reloj de simulación se incrementa en exactamente Δt unidades de tiempo (Δt elegido apropiadamente). Cada vez que se incrementa el tiempo se actualizan las variables de estado, si es un modelo de eventos discretos comprobando si algún evento ha ocurrido en ese intervalo de tiempo. Los eventos que hayan podido ocurrir en ese intervalo se considera que ocurren al final de éste, momento en que se actualizan las variables. Este método es oportuno para simulación continua o cuando el momento en que ocurren los eventos es fijo o para animación gráfica. Sin embargo, para otros casos presenta algunos inconvenientes, fundamentalmente, la simultaneidad de eventos cuando más de un evento ocurre en un intervalo y el error que se comete de

redondeos al considerar que los eventos ocurren al final del intervalo y que para disminuirlo hace que se tomen incrementos muy pequeños que ralentizan enormemente la simulación comprobando muchos intervalos en los que no ocurre nada.

- *Incremento al próximo evento (event step)*: sólo es válido para modelos de eventos discretos, donde el reloj de simulación se inicializa a cero y se determinan los instantes en que sucederán los futuros eventos (todos o los más inmediatos que puedan ocurrir). El reloj de simulación se avanza hasta el instante del suceso más inminente de los futuros eventos (el primero de ellos), actualizando en ese instante el estado del sistema dependiendo del evento de que se trate. Este procedimiento tiene como ventajas respecto al anterior que no tiene errores al considerar tiempos exactos de ocurrencia de los eventos y que es más rápido ya que los periodos en que no hay eventos son saltados.

Obsérvese que en ambos casos es necesario tener definidos cuáles son los eventos futuros y en qué instantes se van a dar, al menos los más inmediatos, lo que se verá en el siguiente capítulo.

I.4.2 Mecanismo de transición

Con cualquiera de los métodos propuestos, cuando se produce un avance del reloj de simulación hay que actualizar las variables de estado. Se define el *mecanismo de transición* como el mecanismo que muestra los cambios que se producen en el estado del sistema y que permite actualizar su valor.

En el caso de los modelos continuos, hay que mostrar los cambios de las variables de estado cuando ha pasado un intervalo de tiempo fijo Δt . Por ejemplo, en el caso de querer hacer un modelo de simulación para el modelo de Lotka-Volterra las variables de estado serían

$X(t)$: número individuos presa en el instante t

$Y(t)$: número de individuos depredador en el instante t

y el mecanismo de transición sería de la forma:

$$\begin{aligned} X' &\leftarrow X + (rX(t) - aX(t)Y(t))\Delta t, \\ Y' &\leftarrow Y + (-sY(t) + bX(t)Y(t))\Delta t \end{aligned}$$

actualizando posteriormente $X \leftarrow X'$ $Y \leftarrow Y'$ (una simplificación inmediata es no pasar por las variables X' e Y' , pero se pueden cometer pequeños errores por la actualización de la variable Y con un valor ya actualizado de X)

En el caso de un modelo de eventos discretos, el mecanismo de transición va asociado con cada evento mostrando los cambios que se producen en el estado del sistema cuando se produce ese evento. Por ejemplo, en un sistema de colas con un servidor, la variable de estado es el número de clientes que hay en el centro de servicio y, por lo tanto, los eventos son la llegada de un cliente y el final de un servicio, y el mecanismo de transición se puede definir como

$$N(t) \leftarrow \begin{cases} N(t)+1 & \text{si llegada cliente} \\ N(t)-1 & \text{si final de servicio cliente} \end{cases}$$

siendo $N(t)$ el número de clientes en el sistema en el instante t .

I.5 Organización de un modelo de simulación dinámico

En el desarrollo de un modelo de simulación de eventos discretos hay otros elementos que intervienen, además del estado del sistema y del reloj de simulación:

- *Lista de eventos*: es la lista de instantes en que van a ocurrir los próximos eventos de cada tipo. Esta lista puede ser generada desde el principio y almacenada (o al menos los valores de las variables aleatorias que los determinan) o puede ir generándose a medida que los valores son necesitados. Como ya se ha dicho en el siguiente capítulo se verá cómo obtenerlos.
- *Contadores estadísticos*: son variables utilizadas para almacenar información sobre el comportamiento del sistema y que al final, mediante algún cálculo matemático, darán respuesta al objetivo del estudio.

Además, un modelo de simulación se estructura en distintas partes, cada una de ellas con un cometido:

- *Rutina de tiempo*: determina el siguiente evento y avanza el reloj de simulación al instante en que va a ocurrir
- *Rutina de evento*: actualiza las variables cuando ha ocurrido un evento. Hay una por cada tipo de evento.
- *Generador de informes* o resultados: realiza los cálculos o estimaciones de las características que se desean medir, cuando la simulación acaba.
- *Programa principal*: enlaza todas las rutinas anteriores.

Plasmado en un diagrama de flujo, éste sería el siguiente (donde no hay numeración es que el orden no es relevante):

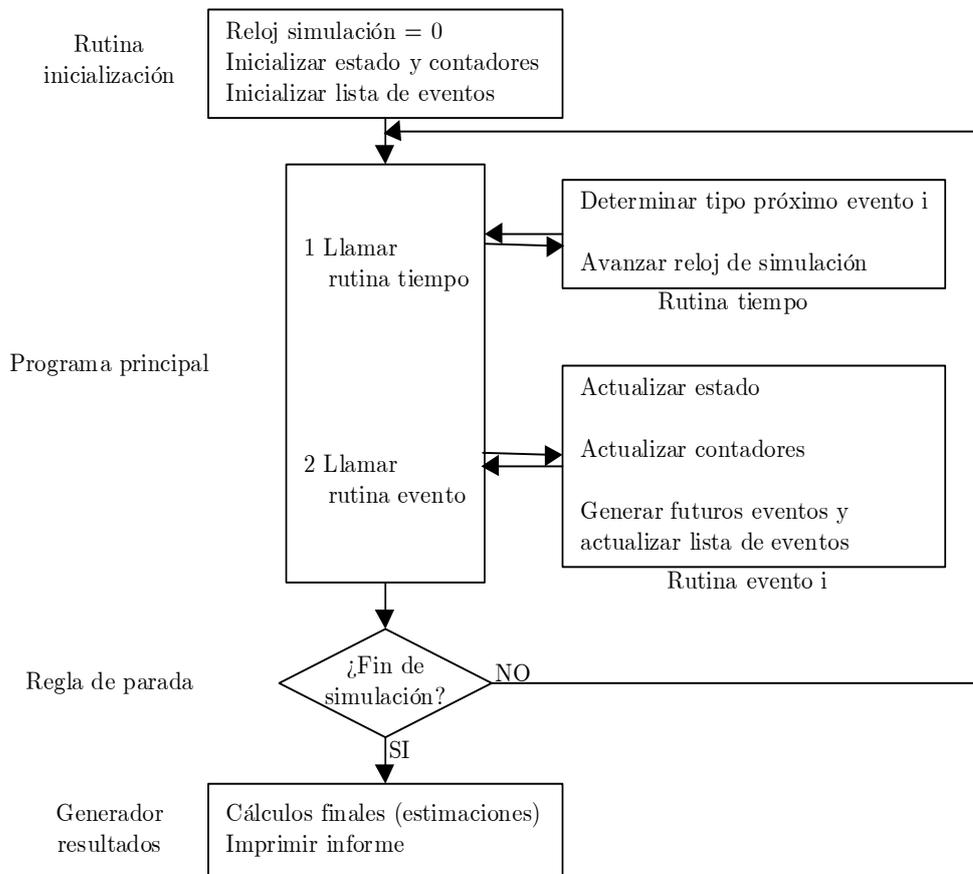


Figura 2. Organización de un modelo de simulación de eventos discretos.

Veamos un ejemplo de un sistema de una línea de espera con un servidor. El objetivo es estimar el número medio de clientes en el sistema.

Se suponen las siguientes hipótesis y datos:

- Tiempos entre llegadas de clientes según distribución F
- Tiempos de servicio según distribución G
- Distribuciones de tiempos independientes entre sí
- T tiempo máximo de simulación

La variable de estado es N el número de clientes en el sistema. Los eventos posibles son:

- Llegada de un cliente al sistema
- Final de servicio de un cliente

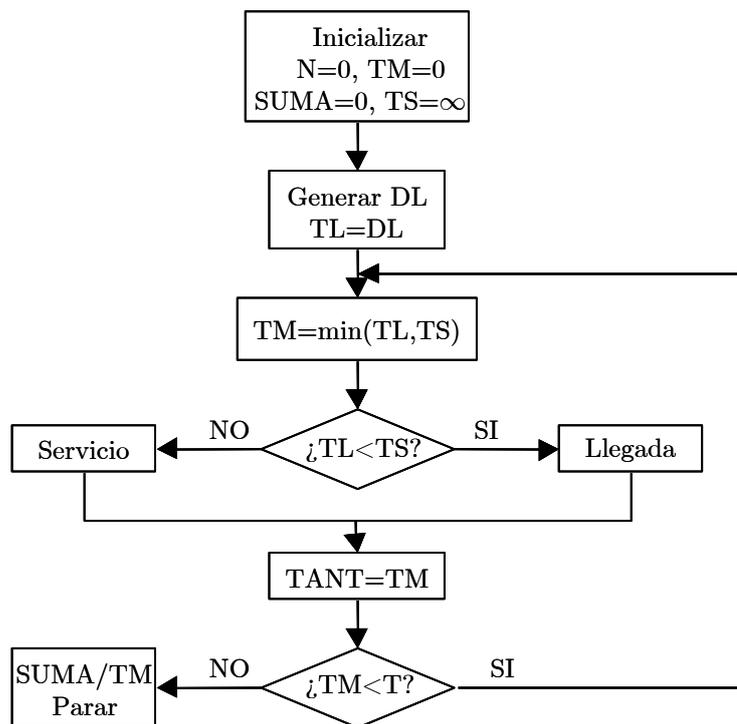
El mecanismo de transición se define como

$$N(t) = \begin{cases} N(t) + 1 & \text{si es llegada de un cliente} \\ N(t) - 1 & \text{si es final de servicio} \end{cases}$$

Otras variables auxiliares son:

- TM reloj de simulación
- DL tiempo entre llegadas $\stackrel{D}{=} F$
- DS tiempo de servicio $\stackrel{D}{=} G$
- TL instante de la próxima llegada
- TS instante del próximo final de servicio
- SUMA contador acumulando suma de productos de clientes en el sistema por tiempo de permanencia
- TANT variable auxiliar (instante de último evento)

El diagrama del modelo se presenta en la figura 3.



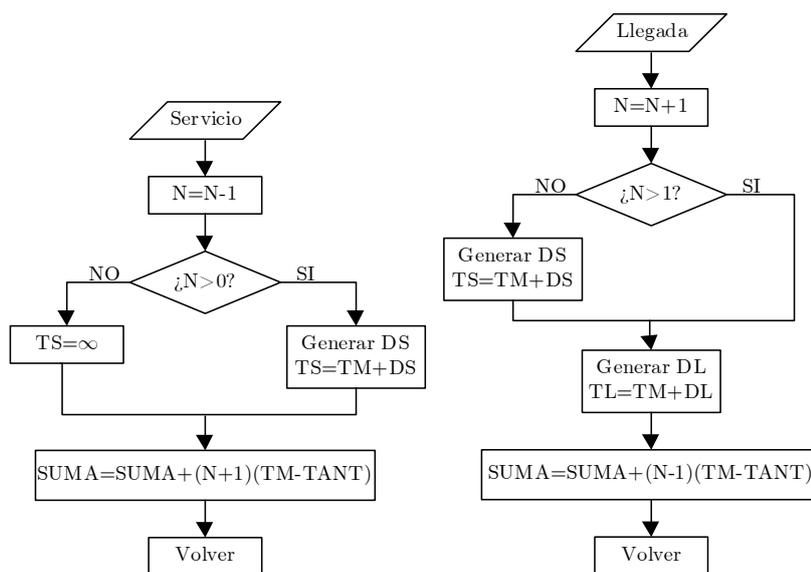


Figura 3. Diagrama del modelo de simulación de una cola con un servidor.

A continuación, se presenta la traza del modelo. La *traza* es una tabla en la que se recogen los valores de las variables que intervienen en el modelo en varias iteraciones. Es útil obtener una traza manualmente para ayudar a obtener el modelo, de modo que se vean las variables que intervienen, los cálculos, etc. Gracias a ella es posible detectar la necesidad de variables auxiliares. Además, es una forma de verificar la programación posterior, de modo que una vez programado el modelo, con los mismos datos de entrada debe ejecutar lo que se recoge en la traza. En caso de que no sea así, puede ser por una programación errónea o incluso un modelado erróneo.

Veamos la ejecución del modelo de simulación. Se obtienen muestras de los tiempos entre llegadas DL dando lugar a estos valores: 3, 2, 5, 1, 2, 6, 6, 2, 8 y de los tiempos de servicio DS que resultan: 4, 1, 3, 1, 3, 2, 3, 5. Se simula durante $T=35$ obteniéndose la siguiente traza.

Nº evento	Reloj Simulación	Tipo evento	N	TL	TS	SUMA
0	0	Inicio	0	3	∞	0
1	3	Llegada	1	5	7	$0+0\cdot3=0$
2	5	Llegada	2	10	7	$0+1\cdot2=2$
3	7	Servicio	1	10	8	$2+2\cdot2=6$

4	8	Servicio	0	10	∞	$6+1 \cdot 1=7$
5	10	Llegada	1	11	14	$7+0=7$
6	11	Llegada	2	13	14	$7+1 \cdot 1=8$
7	13	Llegada	3	19	14	$8+2 \cdot 2=12$
8	14	Servicio	2	19	15	$12+3 \cdot 1=15$
9	15	Servicio	1	19	18	$15+2 \cdot 1=17$
10	18	Servicio	0	19	∞	$17+3 \cdot 1=20$
11	19	Llegada	1	25	21	$20+0=20$
12	21	Servicio	0	25	∞	$20+1 \cdot 2=22$
13	25	Llegada	1	27	28	$22+0=22$
14	27	Llegada	2	35	28	$22+1 \cdot 2=24$
15	28	Servicio	1	35	33	$24+2 \cdot 1=26$
16	33	Servicio	0	35	∞	$26+1 \cdot 5=31$
17	35	Final				$31+0 \cdot 2=31$

Figura 4. Traza del modelo.

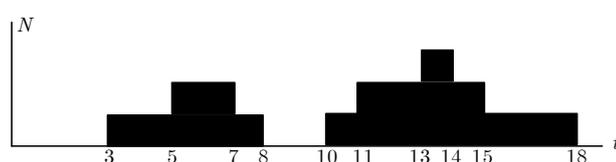


Figura 5. Número de clientes en el sistema a lo largo del tiempo.

Gráficamente el número de clientes en el sistema a lo largo del tiempo se representa en la figura 5. Para el tiempo de simulación $T=35$ el número medio de clientes en el sistema resulta ser $\hat{E}[N] = 31/35 = 0.89$. Si el tiempo de simulación hubiera sido reducido a 18 en lugar de 35 el número medio de clientes en el sistema resultaría $\hat{E}[N] = 20/18 = 1.11$, lo cual pone de manifiesto la importancia de la condición de finalización del proceso de simulación y su posible influencia en los resultados obtenidos.

I.6 Algunas aplicaciones de simulación

I.6.1 Disponibilidad dinámica de un sistema paralelo

Se trata de un sistema formado por varios componentes en paralelo. El sistema está indisponible cuando *todos* los componentes están indisponibles. Se supone que cuando se repara un componente vuelve a funcionar como nuevo.

Se consideran los siguientes datos:

- Distribución de probabilidad de tiempo entre fallos del componente i , $TF_i(t)$
- Distribución de probabilidad de tiempo de reparación del componente i , $TR_i(t)$
- Tiempo máximo de simulación T

La variable de estado de cada componente es $x_i = \begin{cases} 1 & \text{en funcionamiento} \\ 0 & \text{en reparación} \end{cases}$ y la del

sistema $ES = \begin{cases} 1 & \text{sistema en funcionamiento} \\ 0 & \text{sistema en fallo} \end{cases}$

Los eventos que se consideran son:

- Fallo de un componente $x_i = 0$
- Fin de reparación de un componente $x_i = 1$

Otras variables auxiliares son:

- TM reloj de simulación
- TF_i muestra de la distribución de tiempo entre fallos del componente i
- TR_i muestra de la distribución de tiempo de reparación del componente i
- TP_i tiempo del siguiente evento del componente i
- TPS tiempo de parada del sistema
- TI tiempo en que se inició la última parada

El proceso de simulación consta de las siguientes etapas descritas mediante pseudocódigo:

1. Inicialización:

- $TM = 0$
- $TPS = 0$

Definir estado inicial:

- $x_i = 1$
- $ES = 1$

Generar muestra de distribución de tiempo entre fallos TF_i

- $TP_i = TF_i$
2. Rutina de tiempo: Determinar mínimo de TP_i (cuándo), actualizar el tiempo de simulación $TM = TP_i$, determinar el evento de fallo o reparación (qué) y el componente al que afecta (a quién)
 3. Rutina de evento de fallo del componente i
 - $x_i = 0$
 - Si están en fallo el resto de componentes se declara indisponible el sistema $ES = 0$ y se registra el instante de la parada $TI = TM$
 - Generar muestra de distribución de tiempo de reparación TR_i
 - $TP_i = TP_i + TR_i$
 4. Rutina de evento de fin de reparación del componente i
 - $x_i = 1$
 - Si el sistema está indisponible () se contabiliza el tiempo de parada del sistema $TPS = TPS + TM - TI$ y se declara disponible el sistema $ES = 1$
 - Generar muestra de distribución de tiempo entre fallos TF_i
 - $TP_i = TP_i + TF_i$
 5. Regla de parada: Comprobar que $TM < T$, en cuyo caso ir a 2. En caso contrario ir a 6
 6. Final. Índice de disponibilidad $= 1 - TPS / TM$

I.6.2 Fiabilidad estática de un sistema de generación de energía eléctrica

Calcular los índices de fiabilidad *estáticos* de un sistema de generación, denominados probabilidad de pérdida de carga $LOLP = P(D > G)$ y energía esperada no suministrada $EENS = E(D - G)d | D > G$, siendo G la variable aleatoria potencia de generación disponible en el sistema, D la demanda y d la duración de la misma.

Grupo	Potencia [MW]	Tasa de fallo equivalente (EFOR) [%]
1	1000	8
2	1500	10
3	750	7
4	500	6
5	350	5
6	250	3
7	200	2

Cada generador tiene dos estados posibles: disponible e indisponible. La potencia disponible se modela mediante una distribución de Bernoulli, es decir, será igual a la nominal con una probabilidad $(1-EFOR/100)$ e igual a 0 con probabilidad $(EFOR/100)$. Por consiguiente, el número total de estados del sistema será $2^7 = 128$.

Para calcular estos índices de manera exacta para un sistema con un número reducido de generadores se construye la *tabla de estados* del sistema y para cada uno se evalúa la potencia disponible, la potencia deficitaria con respecto a la demanda y la probabilidad de ocurrencia. Luego, la LOLP será simplemente la suma de las probabilidades de los estados donde la potencia deficitaria es estrictamente mayor que 0 y la EENS será la suma ponderada de la potencia deficitaria por su probabilidad. En la tabla se presentan algunos de los 128 estados del sistema (representando con 1 a cada generador disponible y con 0 en caso contrario) y sus parámetros suponiendo una demanda de 3000 MW con una duración de 680 h.

Estado sistema	Pot Disp [MW]	Déficit [MW]	Probab	LOLP	EENS
1111111	4550	0	0.6537	0.0000	0.0
1111110	4350	0	0.0133	0.0000	0.0
1111101	4300	0	0.0202	0.0000	0.0
1111100	4100	0	0.0004	0.0000	0.0
1111011	4200	0	0.0344	0.0000	0.0
1111010	4000	0	0.0007	0.0000	0.0
1011111	3050	0	0.0726	0.0000	0.0
...
1011110	2850	150	0.0015	0.0015	151.2
1011101	2800	200	0.0022	0.0022	305.5

1011100	2600	400	0.0000	0.0000	12.5
1011011	2700	300	0.0038	0.0038	779.8
1011010	2500	500	0.0001	0.0001	26.5
1011001	2450	550	0.0001	0.0001	44.2
1011000	2250	750	0.0000	0.0000	1.2
...
0000110	600	2400	0.0000	0.0000	1.0
0000101	550	2450	0.0000	0.0000	1.6
0000100	350	2650	0.0000	0.0000	0.0
0000011	450	2550	0.0000	0.0000	2.8
0000010	250	2750	0.0000	0.0000	0.1
0000001	200	2800	0.0000	0.0000	0.1
0000000	0	3000	0.0000	0.0000	0.0

Para calcular estos mismos índices mediante simulación se muestrea cada variable aleatoria (potencia disponible de cada generador) según su distribución y se calculan las variables potencia disponible del sistema y déficit. La probabilidad de ocurrencia de cada muestra es la inversa del número total de muestras.

I.6.3 Fiabilidad dinámica de un sistema de energía eléctrica

Determinar los índices de fiabilidad LOLP y EENS y el coste medio de la energía mediante simulación *cronológica* durante al menos 156 semanas consecutivas.

Los datos de la demanda de cada día de una semana característica expresados en MW y los del sistema de generación se dan a continuación.

Hora	1	2	3	4	5	6	7	8	9	10	11	12
L	2936	2893	2675	2676	2533	2652	2546	2794	2848	2623	2616	2811
M	2789	2770	2879	2695	2644	2710	2792	2706	2839	2518	2942	2788
X	2971	2875	2538	2893	2695	2837	2716	2606	2702	2893	2738	2598
J	2689	2894	2685	2718	2742	2954	2675	2985	2714	2646	2615	2660
V	2846	2988	2516	2686	2601	2544	2815	2916	2506	2558	2606	2531
S	2967	2633	2906	2582	2778	2518	2859	2798	2904	2541	2867	2818
D	2959	2500	2575	2957	2975	2824	2792	2766	2744	2686	2958	2861

Hora	13	14	15	16	17	18	19	20	21	22	23	24
L	2951	2638	2712	2506	2617	2551	2507	2962	2898	2902	2930	2800
M	2605	2740	2530	2845	2524	2654	2845	2709	2747	2840	2732	2646
X	2638	2757	2607	2534	2949	2734	2692	2792	2621	2720	2739	2866
J	2697	2908	2503	2573	2771	2970	2582	2831	2552	2820	2602	2861
V	2515	2531	2717	2557	2541	2645	2703	2522	2587	2674	2567	2839
S	2641	2944	2554	2510	2963	2859	2851	2867	2638	2707	2879	2652
D	2624	2797	2914	2738	2791	2771	2933	2841	2920	2638	2507	2668

Grupo	Coste variable [€/MWh]	MTTF [h]	MTTR [h]
1	10	115	10
2	15	54	6
3	25	93	7
4	30	141	9
5	35	95	5
6	50	97	3
7	70	98	2

Las distribuciones de probabilidad de fallo de los grupos son normales con media $\alpha = MTTF$ y desviación típica $\sigma = MTTF/5$. Las de reparación son triangulares con media $m = 2MTTR/3$, valor mínimo $a = MTTR/3$ y máximo $b = 2MTTR$. Los grupos se despachan por orden creciente de coste variable. El coste de la ENS se estima en 1 €/kWh. Por razones de soporte de tensión en la red el grupo 6 debe funcionar si está disponible y por limitación en la evacuación de energía no pueden funcionar simultáneamente el 3 y el 5 teniendo prioridad el de menor coste variable. En el instante inicial todos los grupos están disponibles.

En este ejemplo el tipo de simulación más adecuado es el de avance de tiempo fijo con intervalo unitario de 1 hora (aunque no se hayan producido eventos de fallo o reparación de un grupo), ya que la demanda cambia en cada hora y esto influye en el cálculo del coste medio de producción.

I.7 Etapas en el desarrollo de un estudio de simulación

Al llevar a cabo un estudio de simulación hay algunos pasos que se deben seguir para lograr el éxito, entendiéndose éste como un correcto análisis del problema, la obtención de un modelo que sirva para los fines para los que se desarrolla y que sea creíble por parte del destinatario final.

Por otra parte, hay que diferenciar si el fin del estudio es sólo el desarrollo del modelo (obtener un programa que otro usuario va a utilizar finalmente) o si el análisis de resultados forma parte también del estudio.

A continuación, se muestra el caso más general que incluye el análisis de resultados como ayuda en la toma de decisiones.

FORMULAR EL PROBLEMA

En esta fase se plantean los objetivos del estudio, las hipótesis básicas que pueden ser asumidas, los parámetros que intervienen, cuáles son las variables de estado del sistema y por tanto del modelo, etc. Es una fase clave, denominada *fase de especificación*, que, a menudo, es infravalorada en tiempo, ya que aparentemente el problema está formulado, pero, concretar exactamente lo que se quiere obtener y de lo que se dispone, no es tan obvio como pueda parecer.

REUNIR DATOS Y CREAR UN MODELO

En esta fase en primer lugar hay que reunir los datos. Éste puede ser el cuello de botella del proyecto total ya que la recogida de éstos puede ser una tarea larga y tediosa. Una vez obtenidos, mediante un análisis estadístico, hay que modelar la aleatoriedad para ser incluida en el modelo de simulación.

PROGRAMAR EL MODELO

Cuando se conocen los datos de que se dispone y su naturaleza, se desarrolla el modelo en sí mismo. La programación del modelo, como ya se ha comentado, se puede hacer con un lenguaje de programación de propósito general o con un lenguaje específico de simulación (GPSS, AutoMod, Witness, Arena, ...). Las ventajas e inconvenientes de unos y otros serán analizadas en otra sección. En cualquier caso, la decisión ha de ser tomada previamente, ya que influye no sólo en la programación sino también en el modelo desarrollado. Para el primer caso éste será un diagrama de flujo.

VERIFICAR LA PROGRAMACIÓN

Este paso consiste en verificar que lo que se ha programado coincide con lo que se había modelado. Para ello hay que llevar a cabo ejecuciones de prueba y compararlas con el modelo, para ello una traza bien elegida y desarrollada puede ser una valiosa ayuda. En el caso de ser detectados errores, habrá de volver a la fase anterior.

VALIDAR EL MODELO

Consiste en comprobar la validez del modelo diseñado y para ello hay que ejecutar el modelo y comparar con el propio sistema o con soluciones teóricas de casos sencillos que se hallen bien resueltos. La comparación con el propio sistema sugiere comparación de datos reales y datos simulados. Cuando existe aleatoriedad no es fácil hacer tal comparación y un procedimiento habitual para llevarla a cabo consiste en alimentar el modelo con los mismos datos con los que se alimenta al sistema real para obtener los resultados que se van a comparar. Por ejemplo, si es un sistema de colas con un servidor, se le da al modelo los tiempos entre llegadas y de servicio observados y se compara lo que ocurre en el sistema real con esos tiempos con lo que el modelo simula (no se generan aleatoriamente, en este caso para la validación).

DOCUMENTAR EL MODELO O SIMULADOR

Ésta es una etapa fundamental del desarrollo de un modelo para garantizar su amplia difusión. La documentación ha de ser clara, precisa y completa. El manual de usuario debe incluir la especificación técnica funcional, matemática e informática. El propio código debe incluir una buena documentación para facilitar la tarea del mantenimiento. Piénsese que la mayor parte del ciclo de vida de un modelo no está en el desarrollo sino en la fase de uso y mantenimiento.

Hasta aquí, sería el desarrollo del modelo de simulación, es decir, la construcción de una herramienta que simule un sistema. “Sólo” faltaría para dar por terminada esta parte documentar el modelo con instrucciones al usuario, descripción de hipótesis y datos, etc., y la puesta en servicio para otros usuarios, si es el caso, que sean, al fin y al cabo, los usuarios finales. Esa fase de finalización, puede ser también muy larga y costosa, dependiendo de dónde haya de ser instalado, de la documentación que haya que entregar, de la formación que haya que dar a los usuarios, etc.

Sea quien sea el usuario final (es decir, tanto si es quien ha desarrollado el modelo como si no), el modelo se usará para ayudar en la toma de decisiones y para ello habrá que llevar a cabo ejecuciones con distintas configuraciones, etc. También para esta parte, existe una metodología con el fin de hacer creíbles los resultados obtenidos y obtener la mayor información posible de forma eficaz.

DISEÑAR EL EXPERIMENTO

Hay que diseñar las estrategias a evaluar, las pruebas que se van a llevar a cabo, el número de simulaciones de cada una de ellas, etc. Un buen diseño de experimentos puede ser fundamental para obtener la información eficientemente. También las técnicas de reducción de la varianza que permiten obtener estimadores más precisos con el mismo esfuerzo computacional, o igual de precisos con menor esfuerzo computacional, son una buena ayuda para resolver el problema de forma eficiente. Algunas de estas técnicas están especialmente diseñadas para simulación (variables antitéticas, variables de control, muestreo correlado, ...) y otras provienen de la teoría general de muestreo (muestreo estratificado, muestreo por importancia, ...).

En esta fase también hay que tener en cuenta los periodos transitorios a la hora de diseñar las pruebas, ya que si nuestro objetivo es obtener medidas de un sistema en estado estacionario, habrá que determinar la longitud del periodo transitorio (periodo de tiempo hasta que se puede considerar que el régimen de funcionamiento del sistema es estacionario o permanente) y utilizar procedimientos para eliminar o atenuar su influencia en las medidas que se pretenden obtener.

LLEVAR A CABO LAS EJECUCIONES DE SIMULACIÓN

Se ejecutan todas las pruebas que han sido diseñadas en la fase anterior.

ANALIZAR LOS RESULTADOS

Hay que tener muy en cuenta que lo que se tiene de cada ejecución es una muestra simulada y, por lo tanto, hay que recurrir al análisis estadístico para sacar las conclusiones, igual que si la muestra no fuera simulada.

DECIDIR SI DAR POR TERMINADA LA SIMULACIÓN

A la vista de los resultados, hay que decidir si dar por terminada la simulación y pasar a la última fase del estudio o, por el contrario, si se requieren nuevas pruebas, en cuyo caso habría que volver al paso de diseñar un nuevo experimento.

DOCUMENTAR Y ORGANIZAR LAS EJECUCIONES

Es importante recopilar y mostrar bien la información obtenida de las simulaciones, con el fin de hacer creíbles las conclusiones y decisiones que se propongan. Para ello una buena documentación es un punto fundamental que forma parte del propio estudio.

II Modelado de la aleatoriedad en simulación

Como ya se ha comentado, los modelos de simulación habitualmente incluyen aleatoriedad, hasta el punto de que en muchos casos cuando el modelo es determinista se considera un caso particular de un modelo estocástico más general. Por tanto, es necesario modelar correctamente la aleatoriedad incluida y disponer de procedimientos rápidos y eficientes para generar valores de variables aleatorias con una distribución determinada y conocida.

Este capítulo incluye una breve revisión estadística de los conceptos básicos utilizados en el resto del documento, una sección dedicada a identificar modelos para representar la aleatoriedad a partir de datos conocidos, lo que se ha llamado identificación de patrones y, por último, una sección dedicada a cómo simular esos modelos, es decir, a cómo generar variables aleatorias para ser utilizadas en el proceso de simulación.

II.1 Revisión de probabilidad y estadística

Esta revisión estadística comprende conocimientos estadísticos de aplicación práctica en simulación. Se inicia con una sección dedicada a una revisión de conceptos básicos de probabilidad, después una revisión de estimación puntual y por intervalos y una final de contrastes de hipótesis.

II.1.1 Variables aleatorias y sus propiedades

Un *experimento aleatorio* es un proceso cuyo resultado no se conoce con exactitud. El conjunto de resultados posibles de dicho experimento es lo que se conoce como *espacio muestral*, S . Los resultados concretos de realizar el experimento se denominan *muestras*. Un ejemplo puede ser el resultado de tirar un dado con seis caras. Su espacio muestral sería $S = \{1, 2, 3, 4, 5, 6\}$.

Sin embargo, no todos los experimentos aleatorios tienen resultados numéricos; por ejemplo, al tirar una moneda el espacio muestral sería $S = \{cara, cruz\}$.

Una *variable aleatoria* X es una función que asigna un valor a cada resultado del experimento. Así por ejemplo, si al resultado del lanzamiento de una moneda le asignamos el valor 1 si es cara y -1 si es cruz, tendremos una variable aleatoria cuyos posibles valores o *soporte* es el conjunto $\{-1, 1\}$. Las variables aleatorias son de gran importancia pues trasladan

la probabilidad de un espacio no numérico a uno numérico con todas las ventajas que ello conlleva, entre otras las caracterizaciones y propiedades que se verán a continuación.

Habitualmente se nota a las variables aleatorias con letras mayúsculas y a los valores muestrales de dicha variable con letras minúsculas. Así por ejemplo se habla de la variable X y de las muestras x_i .

Para caracterizar la distribución de probabilidad de una variable aleatoria hay distintas formas según sea ésta, sin embargo una función que puede ser definida de forma general para caracterizar la distribución de una variable aleatoria es la función de distribución.

La *función de distribución* de una variable aleatoria, o también conocida como función de distribución acumulada, $F(x)$, se define como la función que asigna a cada punto la probabilidad de que X tome un valor igual o inferior a x , es decir, $F(x) = P((-\infty, x]) = P(X \leq x)$.

Las propiedades de esta función de distribución son las siguientes:

- Su valor está comprendido entre 0 y 1, es decir, $0 \leq F(x) \leq 1$
- Es no decreciente
- En los límites, se cumple $\lim_{x \rightarrow +\infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$

Las variables aleatorias que pueden ser de nuestro interés, se clasifican en: discretas, absolutamente continuas (en general, las llamaremos continuas) y mixtas. Las que no corresponden a ninguno de estos grupos exceden al objetivo de esta revisión básica.

Se dice que una variable aleatoria es *discreta* cuando puede tomar una cantidad numerable¹ de valores.

Estas variables son caracterizadas, además de por la función de distribución, por la *función de masa de probabilidad*, que es una función que asigna a los posibles valores su probabilidad y al resto 0. Los valores de esta función han de estar entre 0 y 1 ya que son probabilidades y la suma de todos ha de ser 1.

¹ Es decir, una cantidad finita o infinita numerable (los números naturales, los enteros, los racionales,...)

La función de distribución $F(x)$ de este tipo de variables se calcula como la suma de la función de probabilidad para valores iguales o inferiores al valor x , y resulta por lo tanto una función escalonada:

$$F(x) = \sum_{x_i \leq x} P(x = x_i) ; -\infty < x < +\infty$$

Se dice que una variable aleatoria es *absolutamente continua* cuando existe una función, denominada *función de densidad*, $f(x)$, tal que su integral para un conjunto I determina la probabilidad de ocurrencia de dicho conjunto.

$$P(x \in I) = \int_I f(x) dx$$

Se ha de tener en cuenta que la integral de la función de densidad en todo el espacio ha de ser la unidad.

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

En este tipo de variables la probabilidad de ocurrencia de un valor concreto es cero ya que la integral en un punto es nula. Un concepto físico que se asemeja a lo que representa la función de densidad es precisamente la densidad de un cuerpo, en este caso referido a probabilidades.

La función de distribución en este caso se calcula como la integral de la función de densidad en el intervalo $[-\infty, x]$, y de aquí que su derivada constituye la función de densidad asociada a la variable x . La función de distribución en este caso es continua en todos los puntos, y es derivable en todos excepto a lo sumo en una cantidad numerable de ellos.

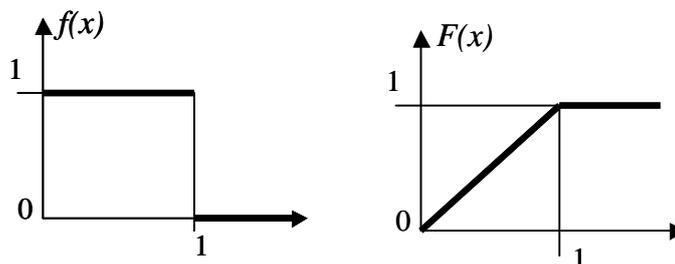
$$F(x) = P(x \in [-\infty, x]) = \int_{-\infty}^x f(y) dy ; -\infty \leq x \leq \infty$$

$$F'(x) = f(x)$$

Así por ejemplo para una variable aleatoria uniforme para el intervalo $[0, 1]$ se tiene que:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{resto} \end{cases}$$

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



Las variables aleatorias *mixtas* son variables que en algunos intervalos se comportan como variables absolutamente continuas, pero que además concentran probabilidad en algún punto. Para ellas no existe otra caracterización teórica válida que no sea la función de distribución, que es continua en todos los puntos excepto en los que acumulan probabilidad que tiene un salto.

Aunque las definiciones que se han dado son válidas en general, hay algunas cuestiones específicas cuando las variables están definidas no en \mathbb{R} , sino en \mathbb{R}^n , siendo denominadas *variables multidimensionales*. Las siguientes definiciones son las equivalentes a las anteriores pero en este contexto, en el que una variable aleatoria es de la forma (X_1, \dots, X_n) :

- *Función de distribución conjunta:*

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

- *Función de densidad conjunta* (en su caso):

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \quad / \quad P(B) = \int_B f(\vec{x}) d\vec{x}$$

- *Función de masa de probabilidad conjunta* (en su caso):

$$P_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \quad / \quad P(B) = \sum_{\vec{x} \in B} p(\vec{x})$$

- *Funciones de densidad marginales* (en su caso):

$$f_{X_i}(x_i) = \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

- *Funciones de masa de probabilidad marginales* (en su caso):

$$P_{X_i}(x_i) = \sum_{x_1} \dots \sum_{x_{i-1}} \sum_{x_{i+1}} \dots \sum_{x_n} P_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$$

Dentro del tratamiento conjunto de variables aleatorias, existe un concepto que es de máxima importancia: el concepto de *independencia*. Se dice que un conjunto de variables

aleatorias es *independiente* si y sólo si la función de distribución conjunta es producto de las funciones de distribución marginales. Esta caracterización para variables absolutamente continuas puede darse diciendo que la función de densidad conjunta ha de ser el producto de las funciones de densidad marginales, y para variables discretas que la función de masa conjunta es el producto de las funciones de masa marginales.

Obsérvese que el concepto de independencia estadística que se acaba de definir no implica que no haya relación alguna entre las variables, sino que el conocimiento de una no modifica la distribución de probabilidad de las otras. Por ejemplo, sea el experimento sacar una carta de una baraja española y observar el palo y número de la carta que ha salido. Así sea la variable X el palo (numerando 1 a copas, 2 a oros, 3 a espadas y 4 a bastos), y la variable Y el número (sota es 8, caballo 9 y rey 10). La función de masa de ambas variables conjuntamente es $p(x, y) = 1/40, \forall x \in \{1, 2, 3, 4\} \forall y \in \{1, \dots, 10\}$. Las funciones de masa marginales son $p(x) = 1/4, \forall x \in \{1, 2, 3, 4\}$ y $p(y) = 1/10, \forall y \in \{1, \dots, 10\}$. Evidentemente, la función conjunta es producto de las marginales, luego, se puede decir que ambas variables son estadísticamente independientes, es decir, si dada una carta se sabe cuál es el palo no modifica la probabilidad de cuál será su número.

Es importante tener presente este concepto de independencia pues en simulación, y en general en probabilidad, se maneja muy habitualmente.

A continuación se presenta una revisión de las características que se consideran más relevantes de una variable aleatoria.

La *esperanza* o *valor esperado* o *media* de una variable X , denotada por μ o $E(X)$, es una medida central de comportamiento de la variable que representa su centro de masas. Se calcula de manera distinta dependiendo de si se trata de una variable discreta o continua.

$$E[X] = \mu = \begin{cases} \sum_{x_i} x_i P(x_i) & \text{siendo } X_i \text{ variable discreta} \\ \int_{\mathbb{R}} xf(x)dx & \text{siendo } X_i \text{ variable continua} \end{cases}$$

La esperanza por su definición es un operador lineal, es decir, $E(\sum_{i=1}^n c_i X_i) = \sum_{i=1}^n c_i E(X_i)$.

Por otra parte, obsérvese que la esperanza de una variable puede ser un valor que no pertenezca al soporte, especialmente en variables aleatorias discretas, aunque siempre se encontrará entre el mínimo y el máximo de los posibles valores.

Otra medida central de una variable aleatoria es la *mediana*. Se nota por $x_{0.5}$ y se define como el menor valor de la variable para el cual $F_X(x) \geq 0.5$. Cuando la variable tiene asimetría es habitual dar este valor de medida central, ya que la media se suele ver muy afectada por los valores más extremos.

La *varianza* de una variable aleatoria es una medida para determinar el nivel de dispersión que tienen los valores de la variable respecto de su media. Se nota por $\text{var}(X)$ o σ^2 , y se define como la media de las desviaciones a la esperanza al cuadrado. Así su definición y un método alternativo de cálculo son:

$$\sigma^2 = E[(X_i - \mu_i)^2] = E(X_i^2) - \mu_i^2$$

Por definición, la varianza es siempre no negativa y es un operador cuadrático, de modo que $\text{var}(cX) = c^2 \text{var}(X)$. Por otra parte, para transformaciones lineales, sólo en el caso en que las variables sean independientes o incorreladas (término que se explica a continuación) se cumple que $\text{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{var}(X_i)$.

Obsérvese que la varianza va en unidades al cuadrado; para dar una medida relativa a las unidades que se están manejando se utiliza la *desviación típica o estándar*, que se define como la raíz cuadrada de la varianza de la variable y se denota por σ .

Calculando la media y la varianza de una variable aleatoria uniforme en el intervalo $[0, 1]$ se tiene:

$$\mu = \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

$$\sigma^2 = \int_0^1 x^2 dx - \left(\frac{1}{2} \right)^2 = \left[\frac{x^3}{3} \right]_0^1 - \frac{1}{4} = \frac{1}{12}$$

Una medida de la dependencia lineal entre dos variables X_i y X_j viene dada por la *covarianza* notada por C_{ij} o $\text{cov}(X_i, X_j)$. Se define como la esperanza o media de la multiplicación de las desviaciones de la variable X_i con respecto a su media μ_i por las desviaciones de la variable X_j respecto de su media μ_j . Así la definición y un método alternativo de cálculo son las siguientes:

$$\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_i X_j) - \mu_i \mu_j$$

Las covarianzas son medidas simétricas, es decir, $C_{ij} = C_{ji}$. Si $i = j$, entonces la covarianza es igual a la varianza de la variable. Si el valor de la covarianza es positivo, ello implica que X_i y X_j están correladas positivamente, es decir, si el valor que se tiene de X_i es mayor que su media μ_i , entonces el valor que se espera para X_j será mayor que su media μ_j . Por el contrario si el valor de la covarianza es negativo, ello implica que X_i y X_j están correladas negativamente, es decir, si el valor que se tiene de X_i es mayor que su media μ_i , entonces el valor que se espera para X_j será menor que su media μ_j .

La covarianza tiene el inconveniente de que es una medida dimensional que no nos permite cuantificar la correlación entre variables de una forma directa. Para obtener una medida adimensional de dicha correlación se utiliza el *coeficiente de correlación lineal*, ρ_{ij} . Dicho factor adquiere un valor entre -1 y 1 y se calcula como cociente entre la covarianza entre dos variables y el producto de sus desviaciones típicas respectivas.

$$\rho_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j}; \quad -1 \leq \rho_{ij} \leq 1$$

Como puede observarse lleva el signo de la covarianza, y su interpretación es por lo tanto la misma.

Aún así, se suele utilizar otro factor para medir el grado de relación lineal entre dos variables: el *coeficiente de determinación*, ρ_{ij}^2 . Este coeficiente representa la proporción de la variabilidad (varianza) de una variable que es explicada por la variabilidad de la otra. Se calcula como el cuadrado del coeficiente de correlación lineal, con lo cual lo que no indica es el signo de la relación pero su interpretación es mucho más intuitiva.

Dos variables se dicen *incorreladas* si el coeficiente de correlación lineal (o el de determinación o la covarianza) es cero. Si dos variables son independientes entonces son incorreladas, al revés no es siempre cierto, excepto en distribuciones normales.

II.1.2 Procesos estocásticos

Un proceso estocástico, $\{X_t\}_{t \in T}$, es un conjunto de variables aleatorias definidas todas sobre un mismo espacio de probabilidad (intuitivamente se diría que miden lo mismo pero en distinto instante de tiempo o punto espacial). Se denomina *espacio de estados* al conjunto de posibles valores de las variables.

Al conjunto T se le denomina *conjunto de índices*, y si se trata de un conjunto numerable de \mathbb{R} , el proceso es denominado *proceso en tiempo discreto*, mientras que si no es un conjunto numerable se denomina *proceso en tiempo continuo*.

Un proceso en tiempo discreto se dice que es *estacionario en covarianza* o *estacionario de segundo orden* si verifica que la media y varianza de todas las variables es la misma y la covarianza entre dos variables del proceso solamente depende de la diferencia entre sus índices y no de sus valores absolutos:

- $\mu_i = \mu ; \forall i$
- $\sigma_i^2 = \sigma^2 ; \forall i$
- $\text{cov}(X_i, X_{i+j}) = \text{cov}(X_1, X_{1+j}) = C_j$ independiente de $i ; \forall j$

En este tipo de procesos a C_j se le denomina *autocovarianza de orden j* , y a los coeficientes de correlación se les conoce como *autocorrelaciones* del proceso.

El arranque de cualquier simulación no suele ser *estacionario en covarianza*. Si la simulación se avanza hasta un tiempo K suficientemente alejado es posible que el proceso alcance un régimen estacionario en el cual se cumplan los requisitos anteriormente descritos. El tiempo que transcurre hasta que se alcanza dicho régimen estacionario se conoce como período transitorio o de calentamiento (*warm up*).

Los *procesos en tiempo continuo* también pueden ser estacionarios o no.

Hay procesos estocásticos que no son estacionarios en sí mismos, pero que pueden serlo sus incrementos. Un *proceso markoviano* es un proceso estocástico de incrementos independientes y estacionarios, es decir, tal que

i) $\forall t, a > 0$ la variable aleatoria $X_{t+a} - X_t$ es independiente de $\{X_s / 0 \leq s < t\}$, es decir, independiente de su historia (o que no tiene memoria).

ii) $\forall 0 < t_1 < t_2 < t_3 < t_4 \quad X_{t_2} - X_{t_1} \stackrel{d}{=} X_{t_4} - X_{t_3}$

Respecto a los *procesos markovianos en tiempo continuo* los más conocidos son el Proceso de Poisson (que es un proceso de conteo ya que el espacio de estados es $\mathbb{N} \cup \{0\}$) y el Proceso de Wiener o movimiento browniano (cuyo espacio de estados es \mathbb{R})

Se dice que un proceso estocástico $\{X_t\}_{t \in T}$ es un proceso de Poisson de parámetro λ si $T \subseteq [0, +\infty)$ con las siguientes propiedades

- i) $X_0 = 0$ c.s.
- ii) $\forall t, a > 0$ la variable aleatoria $X_{t+a} - X_t$ es independiente de $\{X_s / 0 \leq s < t\}$ (es decir, que los incrementos son independientes de su historia)
- iii) $\forall t, a > 0$ $X_{t+a} - X_t \stackrel{d}{=} \wp(\lambda a)$

Se dice que un proceso estocástico $\{X_t\}_{t \in I}$ es un proceso de Wiener o movimiento Browniano si si $T \subseteq [0, +\infty)$ con las siguientes propiedades

- i) $X_0 = 0$ c.s.
- ii) $\forall t, a > 0$ la variable aleatoria $X_{t+a} - X_t$ es independiente de $\{X_s / 0 \leq s < t\}$
- iii) $\forall t, a > 0$ $X_{t+a} - X_t \stackrel{d}{=} N(0, \sqrt{a})$

Se puede probar que con probabilidad uno X no es diferenciable en ningún punto del intervalo $[t, t + a]$.

II.1.3 Estimación puntual

Esta sección se dedica a la estimación puntual, es decir, obtener estimadores para parámetros desconocidos de una variable a partir de una muestra. Los estimadores en general se denotan de una forma particular o con el parámetro que se desea estimar con un circunflejo, es decir, de la forma \hat{a} . La estimación en general se hace a partir de una muestra aleatoria simple y así se presentan los estimadores siguientes.

Sea una muestra aleatoria simple, es decir, X_1, X_2, X_3, \dots son v.a.i.i.d. con media μ y varianza σ^2 , se define como *media muestral* a la suma de los valores muestrales de tamaño n dividido por el propio n . Se nota la media muestral para un tamaño n como $\bar{X}(n)$ o $\hat{\mu}$.

$$\bar{X}(n) = \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

La media muestral es un estimador insesgado² o centrado de la media μ de la variable, es decir, $E(\bar{X}(n)) = \mu$, y es en sí una variable aleatoria.

Así por ejemplo, en un sistema de espera de llamadas se producen los siguientes tiempos de espera en segundos {45, 34, 66, 49, 87, 22, 54, 91}. La media muestral de estos tiempos resulta de 56 segundos/llamada. Este valor es una realización de la variable $\bar{X}(n)$.

La media muestral como tal variable tiene por esperanza la media de la variable y por varianza la varianza de la variable partido por el tamaño muestral, es decir, $E[\bar{X}(n)] = \mu$ y $V[\bar{X}(n)] = \sigma^2 / n$ (la varianza lo es sólo si son variables independientes o incorreladas).

Para estimar la varianza de una variable si fuera conocida (no estimada) la esperanza de la variable en cuestión el estimador utilizado es $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$. Sin embargo, no es habitual que se conozca la media con certeza, sino que suele ser estimada con la media muestral. En tal caso, el estimador utilizado es la cuasivarianza.

La *cuasivarianza o varianza muestral* es un estimador insesgado de la varianza de la variable, es decir, $E(S^2(n)) = \sigma^2$. La cuasivarianza para un tamaño n de la muestra se nota por $S^2(n)$ o por $\hat{\sigma}^2$.

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

Así por ejemplo para el conjunto de tiempos de espera la cuasivarianza muestral obtenida, $S^2(n)$, es igual a 585.714 segundos² (calculado como 4100/7).

Por lo tanto, un estimador de la varianza de la media muestral, $\text{var}[\bar{X}(n)]$, será el cociente entre la cuasivarianza muestral y el número de muestras.

$$\text{var}[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^n (x_i - \bar{X}(n))^2}{n(n-1)}$$

² Un estimador de un parámetro es insesgado o centrado cuando el valor esperado de dicho estimador es el propio parámetro que se desea estimar

Así por ejemplo para la media muestral de 56 segundos/llamada obtenida para los tiempos de espera se obtiene un estimador de la varianza de la media muestral de 73.214 segundos² (calculada como 585.714/8).

II.1.4 Estimación por intervalos de confianza

La estimación puntual en general es insuficiente como único resultado cuando se desea estimar un parámetro ya que no proporciona ninguna información acerca de la precisión del resultado. Por otra parte, cuando se hace una estimación puntual se puede tener prácticamente la seguridad de que el valor estimado no es el valor real. Por lo tanto, junto a una estimación puntual siempre se debe hacer una estimación por intervalo.

Existen distintos métodos para obtener intervalos de confianza, siendo el más conocido el de la cantidad pivotal. Sin embargo en esta sección de repaso no se va a entrar en el procedimiento para obtener intervalos de confianza sino que se van a presentar sólo los utilizados para la media que son los más habituales para analizar los resultados de simulación.

Si la varianza de la variable es conocida (que no estimada), lo que ocurre en muy pocas ocasiones, el intervalo de confianza $1 - \alpha$ para la media es

$$\left(\bar{X}(n) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}}, \bar{X}(n) + z_{\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} \right)$$

donde $z_{\alpha/2}$ denota el valor que deja a su derecha un probabilidad $\alpha/2$ en una distribución normal estándar (media 0 y desviación 1).

Si la varianza de la variable es desconocida y estimada mediante la cuasivarianza, que suele ser lo habitual, el intervalo de confianza $1 - \alpha$ para la media es

$$\left(\bar{X}(n) - t_{n-1, \alpha/2} \cdot \sqrt{\frac{S^2(n)}{n}}, \bar{X}(n) + t_{n-1, \alpha/2} \cdot \sqrt{\frac{S^2(n)}{n}} \right)$$

donde $t_{n-1, \alpha/2}$ denota el valor que deja a su derecha una probabilidad $\alpha/2$ en una distribución T de Student con $n - 1$ grados de libertad. Obsérvese que para valores altos de n la normal y la T de Student son prácticamente iguales.

La interpretación que se ha de dar a los intervalos de confianza según están contruidos es que si se obtuvieran distintos intervalos en $100(1 - \alpha)$ % de ellos se encontraría el parámetro μ , lo que no quiere decir que en un caso concreto se pueda asegurar que se encuentra el parámetro.

Obsérvese que para reducir el tamaño del intervalo de confianza es necesario incrementar el número de muestras n obtenidas o disminuir la cuasivarianza. La variación del tamaño del intervalo con el número de muestras es cuadrática. Así, en el caso de que se quiera reducir el tamaño a la mitad es necesario incrementar en cuatro veces el número de muestras obtenidas.

Así por ejemplo para los tiempos de espera de las llamadas se dispone de los valores necesarios para establecer el intervalo de confianza del 99 % para la media, donde $\bar{X}(n)=56$ segundos/llamada. Mirando en las tablas de la t de *Student* se busca el valor que deje una probabilidad de 0.005 de probabilidad a su derecha. Este valor en tablas resulta 3.499. El intervalo de confianza resultante se indica a continuación. Este intervalo es muy grande debido al poco número de muestras utilizado, al valor de la cuasivarianza y a la elevada confianza que se pide en el intervalo.

$$\left[56 - 3.499 \cdot \sqrt{\frac{585.714}{8}}, 56 + 3.499 \cdot \sqrt{\frac{585.714}{8}} \right] = [26.061, 85.939]$$

II.1.5 Contrastes o tests de hipótesis

Los contrastes de hipótesis se plantean cuando se quiere contrastar alguna hipótesis que se tenga sobre la variable o la muestra. Hay contrastes sobre parámetros, sobre la propia distribución, sobre la aleatoriedad de la muestra, etc.

El planteamiento básico que se va a presentar (no es el único) es siempre el mismo. Se formula una hipótesis, denominada *hipótesis nula* y denotada por H_0 , y lo contrario a ella será la *hipótesis alternativa* que se denota por H_1 .

A partir de la muestra se obtiene un estadístico, tal que si la hipótesis nula fuera cierta se sabría su distribución. Así si el valor del estadístico para la muestra concreta es un valor “raro” para esa distribución se rechaza la hipótesis nula, mientras que si no es raro, no se puede rechazar.

Básicamente, la idea es esa, pero hay varias cosas que puntualizar.

La primera puntualización es que los tests no son simétricos, es decir, no tratan por igual a la hipótesis nula y a la alternativa. Al realizar un test hay dos errores que se pueden cometer: rechazar la hipótesis nula siendo verdadera o rechazar la hipótesis alternativa siendo verdadera. Al primer error se le denomina *Error de tipo I* y al segundo *Error de tipo II*. Los tests se realizan a un nivel de confianza α que representa una cota para la probabilidad del error de tipo

I, es decir, $P(\text{Error tipo I}) \leq \alpha$, mientras que no se tiene ningún control acerca de la probabilidad del error de tipo II³. Por lo tanto, fijado un valor de α pequeño, se sabe que la probabilidad de cometer el error de rechazar H_0 siendo cierta es muy pequeña, mientras que no se sabe nada de la otra. Ésta es la razón por la que habitualmente en el resultado de un contraste si se rechaza la hipótesis nula se hace con seguridad, mientras que cuando no se rechaza se suele decir que no hay evidencia estadística para rechazarla. Es importante tener esto en cuenta cuando se diseña un contraste.

La siguiente puntualización es acerca de lo “raro” de una distribución. En general, se considera “raro” la zona de las colas de las distribuciones, y se delimita lo raro por el *nivel de significación* α . Veamos un ejemplo de contraste completo para definir conceptos.

Supóngase que se quiere contrastar un valor μ_0 para la media de una variable. Entonces el planteamiento del contraste sería

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

En este caso, no habría duda, ya que los contrastes paramétricos obligan a que la igualdad siempre se ponga en la hipótesis nula.

El estadístico que se utiliza será:

$$T = \frac{\bar{X}(n) - \mu_0}{\sqrt{S^2(n)/n}}$$

ya que este valor si la media es realmente μ_0 debería seguir una distribución T de Student con $n-1$ grados de libertad. Si el valor que resulta a partir de una muestra es muy grande en valor absoluto resulta un valor “raro”. Para que además sea cierto que la probabilidad de error de tipo I está acotada por α hay que definir una región, denominada *región crítica*, que siendo cierta la hipótesis nula tenga esa probabilidad, es decir, una región de la T de Student de probabilidad α y que sean valores raros de esta distribución: los intervalos $(-\infty, -t_{n-1, \alpha/2}) \cup (t_{n-1, \alpha/2}, \infty)$. Así si

³ Los tests que se utilizan suelen ser de máxima potencia (mínima probabilidad error de tipo II) para un valor de α , pero aún así queda sin acotar y con un valor siempre desconocido. Aunque una cosa es cierta, cuanto menor sea la cota del error de tipo I (α), mayor es la probabilidad de cometer el error de tipo II:

el estadístico toma valores en esa región la hipótesis nula es rechazada, si no es así, se dice que no puede ser rechazada.

Obviamente, cada contraste vendrá definido por un estadístico que calcular y una región crítica, y al llevarlo a cabo habrá que definir cuál es la hipótesis nula y la alternativa y el nivel de significación teniendo en cuenta lo que suponen al llevarlo a cabo.

Por otra parte, la decisión del nivel de significación puede ser un poco drástica y además contrastes para distintas hipótesis nulas no pueden ser comparados si nos sale por ejemplo no rechazar la hipótesis, así que se impone tener una medida del grado de seguridad con que se acepta o rechaza una hipótesis. Esta medida es lo que se denomina *p-valor*, a veces también *z-value*, etc. Este valor representa el nivel de significación para el que con esa muestra sería rechazada la hipótesis nula, es decir, el tamaño que debería tener la región crítica para que fuera rechazada la hipótesis nula, o lo que es lo mismo, la probabilidad que habría que admitirse de cometer el error de tipo I para rechazar la hipótesis nula.

¿Cómo trabajar con el *p-valor*? Una forma es sustituir la idea de ver si el estadístico está en la región crítica, y en lugar de ello, comparar el *p-valor* con el nivel de significación: si el *p-valor* es menor que α se rechaza la hipótesis nula, si no, no.

Otra posibilidad, y es la que justifica en mayor medida el uso del *p-valor*, es cuando se hacen varios contrastes con distintas hipótesis pero hay que elegir una de ellas. Por ejemplo, en un contraste de bondad de ajuste la hipótesis nula es que la variable sigue una determinada distribución frente a que no lo hace. Supóngase que se busca una distribución entre exponencial, gamma y weibull. Los contrastes obligan a hacer uno cada vez, y supóngase que para los tres se obtiene que no se puede rechazar que la variable siga esa distribución. ¿Qué decisión tomar? O lo que es lo mismo, cuando no rechazo la hipótesis nula, ¿en qué caso lo hago con mayor seguridad? La medida para comparar es justo el *p-valor*, en aquél contraste en que el *p-valor* sea mayor será en el que estoy aceptando la hipótesis nula con mayor seguridad.

II.2 Identificación de patrones

Para generar valores de las variables de entrada de un modelo y poder desarrollar la simulación hay que modelar la aleatoriedad que rige a éstas variables. Este modelado se puede hacer mediante dos aproximaciones:

1. *Ajustar una distribución teórica a los datos*: lo que supone buscar una distribución teórica que sea acorde a los datos disponibles, contrastando la bondad del ajuste para confiar en que

el modelo es el apropiado, y usar esta distribución para generar valores luego de esa variable.

2. *Definir la distribución empírica de los datos:* de modo que la distribución de los datos es utilizada directamente para generar valores de la variable posteriormente.

La aproximación 2 debe usarse sólo si no se puede usar la 1, por las siguientes razones:

- Los datos son aleatorios, diferentes datos producen diferentes distribuciones empíricas. Es decir, el método de la distribución empírica está muy sujeto a fluctuaciones en los datos.
- La distribución empírica no permite generar valores fuera del rango en que se mueven los datos con la que se ha obtenido. Habitualmente, no hay un rango claramente establecido y por lo tanto, usar este método fuerza a que los valores no puedan ser ni menores ni mayores que los ya obtenidos, es una limitación, en general, poco deseable.
- Podemos tener razones para poner un modelo teórico por las propiedades que el modelo seleccionado pueda tener y que son conocidas de antemano.

En lo que sigue se supondrá que se parte de una muestra aleatoria simple, es decir, observaciones independientes de la variable a modelar obtenidas en las mismas condiciones.

II.2.1 Distribuciones empíricas

En esta sección se muestra como obtener la distribución empírica a partir de unos datos. La primera distinción es si se considera que la variable aleatoria es continua o discreta.

Variables aleatorias discretas

Si la variable aleatoria es discreta, la distribución empírica se obtiene asignando a los valores observados la frecuencia relativa con la que aparecen en la muestra. Este procedimiento es muy habitual, por no decir el que más se utiliza, para variables aleatorias discretas, especialmente cuando no existen fundamentos para suponer que la distribución pueda seguir un modelo concreto del tipo Binomial, Poisson o relacionadas.

Variables aleatorias continuas

En el caso de las variables aleatorias continuas cabe destacar a su vez dos casos, que las observaciones se den agrupadas por intervalos o que no sea así.

- *Variables continuas con observaciones no agrupadas:* X_1, \dots, X_n

El procedimiento para obtener la función de distribución empírica es el siguiente:

1) Ordenar los valores de la muestra de menor a mayor: $X_{(1)}, \dots, X_{(n)}$

2) Obtener la función de distribución según la siguiente expresión:

$$F(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{i-1}{n-1} + \frac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & X_{(i)} \leq x < X_{(i+1)} \quad i = 1, \dots, n-1 \\ 1 & x \geq X_{(n)} \end{cases}$$

Por ejemplo, la función de distribución correspondiente a las observaciones 12, 15, 9, 7, 3, 16, sería la correspondiente a la siguiente gráfica:

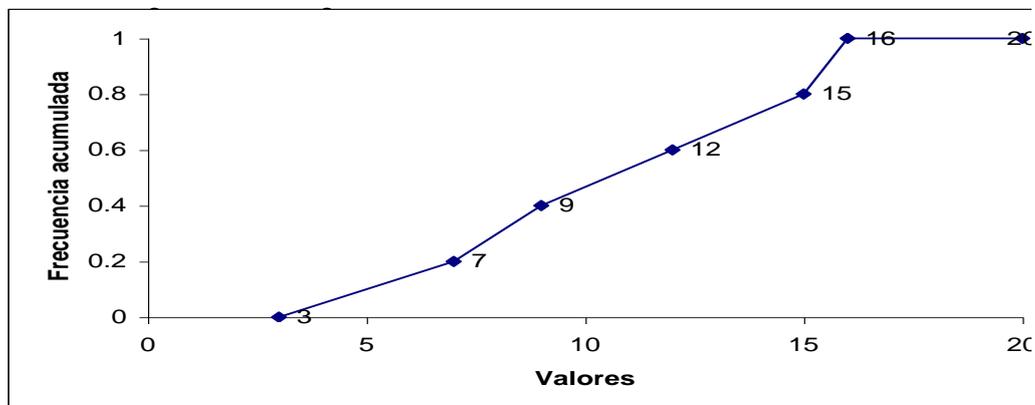


Figura 6. Función de distribución empírica de datos no agrupados

- *Variables continuas con observaciones agrupadas en intervalos:*

Sean n observaciones agrupadas en k intervalos

$[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$, cuyas frecuencias absolutas observadas son n_i de modo que

$$n_1 + n_2 + \dots + n_k = n.$$

La función de distribución empírica, donde $G(a_j)$ denota la frecuencia relativa acumulada,

es decir, $G(a_j) = \frac{\sum_{i=1}^j n_i}{n}$, sería la siguiente:

$$G(x) = \begin{cases} 0 & x < a_0 \\ G(a_{j-1}) + \frac{x - a_{j-1}}{a_j - a_{j-1}} (G(a_j) - G(a_{j-1})) & a_{j-1} \leq x < a_j, j = 1, \dots, k \\ 1 & x \geq a_k \end{cases}$$

Por ejemplo, dada la siguiente tabla de valores, la gráfica de la función de distribución empírica correspondiente es la que se muestra en la figura

[2,4)	[4,8)	[8,10)	[10,15)	$n = 25$
3	10	8	4	

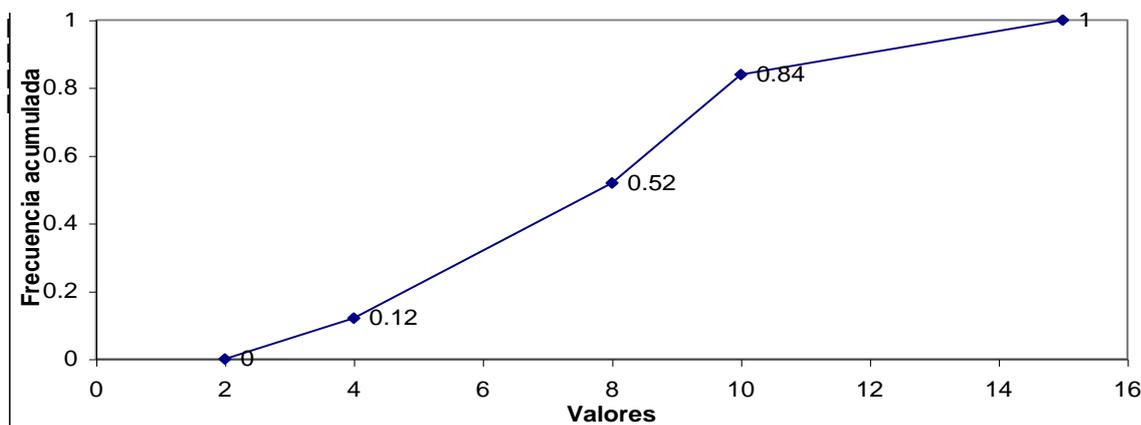


Figura 7. Función de distribución empírica de datos agrupados

II.2.2 Distribuciones teóricas

En esta sección se presenta el procedimiento a seguir para identificar un modelo de distribución teórico apropiado para las observaciones de que se dispone. El procedimiento consta de una serie de pasos que son los siguientes:

- a) Proponer posibles familias de distribuciones a partir de un análisis exploratorio de los datos de tipo descriptivo. Los tres elementos que se utilizan fundamentalmente en este análisis son:
 - Valores característicos de los datos

- Cuantiles y gráficos Box-Plot
 - Histogramas
- b) Estimar los parámetros de las distribuciones propuestas
- c) Determinar cómo de representativas son las distribuciones ajustadas, mediante procedimientos heurísticos y mediante contrastes de hipótesis de bondad de ajuste (Test de la χ^2 y Test de Kolmogorov-Smirnov).

Vamos a desarrollar ahora cada uno de estos pasos con más detalle.

- a) *Proponer familias de distribuciones a partir de un análisis descriptivo.*

Esta fase consiste fundamentalmente en proponer familias de distribuciones que se observe que son apropiadas para los datos a partir del análisis descriptivo de éstos, o lo que ocurre en muchas ocasiones, más que proponer, esta fase permite descartar determinadas distribuciones.

Este análisis puede apoyarse también en determinadas hipótesis que se puedan hacer sobre los datos; por ejemplo, supongamos que nuestros datos son datos de tiempo entre fallos de máquinas y sospechamos o suponemos que estos fallos no están influenciados por el desgaste, sino que son por causas aleatorias, parece razonable plantearse la exponencial como una posible distribución, ya que esta distribución no tiene memoria.

Para poder plantear posibles distribuciones apropiadas para los datos disponibles se deben analizar los valores característicos de los datos, los cuantiles y el gráfico Box-Plot, y el histograma de los datos.

- *Valores característicos de los datos:*

Los valores que suelen ser obtenidos y observados son media, varianza, desviación típica, mediana, asimetría, curtosis, coeficiente de variación, ... La razón es porque hay familias de distribuciones para los que estos valores tienen un determinado valor. Así, por ejemplo, para una distribución normal, la media y la mediana coinciden, y la asimetría y curtosis son cero.

Otro valor característico de una distribución es el coeficiente de variación, es decir, el cociente entre la desviación típica y la media, que es 1, ya que ambos valores son iguales para la exponencial.⁴

⁴ En ocasiones el coeficiente de variación viene dado en porcentaje, siendo entonces 100 el valor característico de una exponencial.

Una matización que realizar es que dado que los valores obtenidos son de la muestra, cuyas observaciones son aleatorias, no se puede pretender que estos valores sean exactos, es decir, aunque los datos provengan de una exponencial, lo habitual es que el coeficiente de variación calculado a partir de los datos no sea 1, pero sí próximo a él.

- *Cuantiles y Box-Plot*

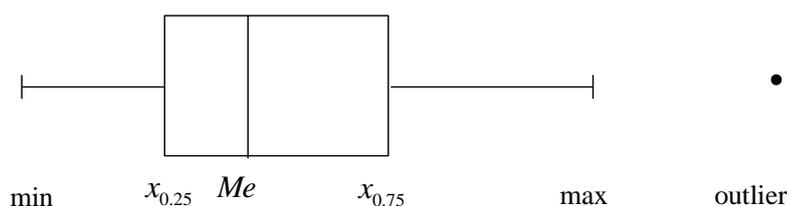
Este análisis suele ser razonable para distribuciones continuas, no para discretas.

Los cuantiles x_q son los valores que dejan a su izquierda una probabilidad q , es decir, un valor tal que $F(x_q) = q$. Es habitual obtener los denominados cuartiles ($x_{0.25}$, $x_{0.5}$ que es la mediana, y $x_{0.75}$) de la muestra y compararlos con los de la distribución propuesta. Esta comparación se suele hacer más que numéricamente por forma, obteniendo el denominado gráfico Box-Plot.

En el gráfico Box-Plot, se representan los cuartiles en una caja central, con una línea en medio que es la mediana; a su vez se representan dos líneas a cada lado hasta el mínimo de los datos la de la izquierda y hasta el máximo la de la derecha, siempre que la longitud de las líneas no supere 1.5 veces el rango intercuartílico; en el caso en que esta cantidad sea superada, los datos que quedan fuera se representan separadamente, siendo denominados *outliers* u observaciones fuera de rango. Los que están distanciados más de 3 veces el rango intercuartílico se denominan *superoutliers*.

Este gráfico se utiliza para ver la asimetría y analizar esos posibles *outliers*. Los *outliers* han de ser estudiados y eliminados del estudio si se concluye que pueden provenir de errores. Sin embargo, ha de tenerse mucho cuidado al tomar la decisión de eliminarlos del análisis. El gráfico Box-Plot considera que son *outliers* para una distribución normal, pero si trabajamos con una distribución asimétrica a la derecha, por ejemplo, lo normal es que aparezcan este tipo de datos sin que sea ningún tipo de error.

Un ejemplo de un gráfico Box-Plot de una distribución asimétrica a la derecha es el que se muestra a continuación



- *Histogramas y diagrama de barras:*

Los histogramas son una representación de los datos provenientes de distribuciones continuas que deben asemejarse a la función de densidad que se proponga. Los diagramas de barras son para datos provenientes de una distribución discreta que deben parecerse a la función de masa de la distribución discreta a proponer.

Los histogramas representan los datos agrupados por intervalos, así a la hora de hacer esta representación, la decisión fundamental es cómo agruparlos. El procedimiento para construir un histograma es el siguiente:

- Dividir el rango en intervalos

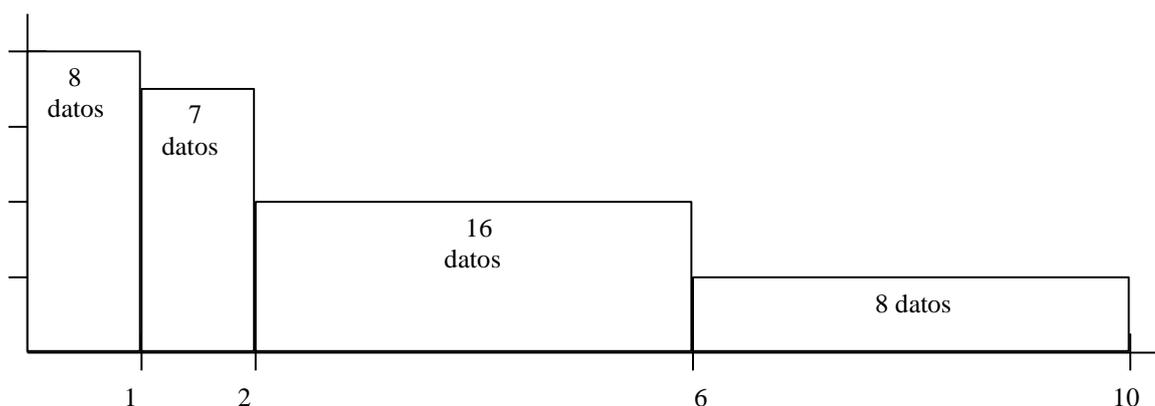
El número de intervalos no es un número fijo sino que se puede probar con varios, teniendo en cuenta que no pueden quedar intervalos de frecuencia nula. En general, los intervalos se suelen hacer todos de igual amplitud, excepto, en todo caso, los extremos. No hay que olvidar que con el histograma pretendemos ver gráficamente si los datos tienen una distribución que se pueda parecer a alguna conocida, así que se debe probar con distinto número de intervalos, y con distintas amplitudes, buscando la semejanza con alguna distribución.

Como orientación para el número de intervalos, se suele probar con números cercanos a la raíz cuadrada del número de datos o a la fórmula de Sturges:

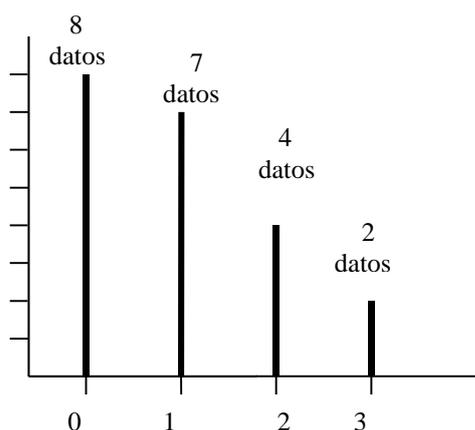
$$N^{\circ} \text{ intervalos} \cong k = \lceil 1 + \log_2 n \rceil = \lceil 1 + 3.322 \log_{10} n \rceil$$

- Levantar sobre cada intervalo un rectángulo de área proporcional a la frecuencia, manteniendo la misma constante de proporcionalidad en todos los intervalos.
- Comparar con las funciones de densidad

Un posible histograma se muestra a continuación. En general, las herramientas informáticas estadísticas construyen los histogramas con toda facilidad, representando incluso la función de densidad de una distribución propuesta a su lado. Sin embargo, aunque se utilicen estas herramientas, la decisión de cómo agrupar los datos para lograr una semejanza si es posible, sigue siendo una tarea a realizar por el usuario.



Para el caso de distribuciones discretas, el diagrama de barras resulta mucho más sencillo, ya que consiste en levantar una barra sobre los valores observados de altura la frecuencia relativa de éstos.



b) Estimar los parámetros de las distribuciones propuestas

Propuestas unas distribuciones, se estiman los parámetros de éstas que hacen que el ajuste a los datos sea el mejor. La estimación se puede hacer mediante distintos métodos de estimación paramétrica, como el método de máxima verosimilitud, el de los momentos, etc. Actualmente, esta fase no se suele llevar a cabo de forma independiente, ya que las herramientas informáticas de estadística realizan la estimación al hacer los contrastes de bondad de ajuste del paso siguiente.

c) Determinar cómo de representativas son las distribuciones ajustadas

Para determinar si son razonables las distribuciones propuestas, el primer paso es comparar gráficamente los histogramas con las funciones de densidad en el caso de distribuciones continuas, y los diagramas de barras con las funciones de masa para el

caso de distribuciones discretas. Este paso a menudo se hace al principio, para hacer la propuesta de distribuciones.

Sin embargo, no es suficiente con ver que gráficamente hay una semejanza sino que además se debe cuantificar la bondad del ajuste para poder comparar cuán bueno es éste con distintas distribuciones, y tomar una decisión acerca de la distribución a seleccionar. Para ello se realizan contrastes de bondad de ajuste con el fin de descartar alguna de las distribuciones propuestas, y para cuantificar la bondad del ajuste mediante el p-valor.

Tests de Bondad de Ajuste

Son contrastes de hipótesis donde el planteamiento es:

Sea X_1, \dots, X_n una muestra aleatoria simple (v.a.i.d.) con distribución F desconocida.

Sea F_0 una distribución particular.

El contraste que se plantea es:
$$\left. \begin{array}{l} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{array} \right\}$$

Observaciones:

1. Por la naturaleza de los contrastes en general, si el resultado es rechazar la hipótesis nula se hará con bastante seguridad; si el resultado es aceptarla se debe decir "no rechazar H_0 ya que no hay evidencia estadística para ello". En general, los tests tienden a aceptar que los datos pueden provenir de la distribución propuesta.
2. Para una cantidad grande de datos es habitual que el resultado sea rechazar H_0 , ya que algunos contrastes son muy sensibles a pequeñas variaciones.
3. Se debe utilizar el p-valor como medida del ajuste:

El p-valor es el nivel de significación para el que se rechazaría H_0 con los datos utilizados para realizar el contraste, es una cota de la probabilidad de cometer el error de tipo I (rechazar H_0 siendo cierta). Cuanto mayor sea el p-valor mejor es el ajuste.

Sin embargo, no se ha de ser ciego a la hora de seleccionar una distribución, ya que puede ser preferible una distribución más sencilla aunque tenga un p-valor peor (siempre que sea razonable) que otra que tenga un mejor valor pero sea muy compleja.

Existen dos tipos de tests de bondad de ajuste que son los más utilizados: el test de la χ^2 y el test de Kolmogorov-Smirnov. En general, estos tests están incluidos en todas las herramientas informáticas de estadística, pero, es recomendable saber en qué consisten con el fin de dar los parámetros adecuados, saber cuándo utilizarlos y saber interpretarlos.

Test de la χ^2 (Pearson, 1900)

La idea básica de este test es agrupar los datos en intervalos y comparar las frecuencias de estos intervalos con las probabilidades que la distribución teórica les asigna, midiendo la distancia entre ambas.

Este test es aplicable tanto a distribuciones discretas como continuas, aunque requiere tener una muestra suficientemente grande ya que se basa en un resultado asintótico.

El procedimiento sería el siguiente:

- a) Dividir el rango de la distribución ajustada en k intervalos adyacentes:

$$[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$$

- b) Sea $N_j =$ número de X_i en intervalo j -ésimo = frecuencia absoluta observada de intervalo $[a_{j-1}, a_j)$

- c) Calcular de la distribución teórica la probabilidad de cada intervalo:

$$p_j = P([a_{j-1}, a_j)), \text{ y a partir de ella la frecuencia esperada como } n \cdot p_j.$$

- d) Calcular:
$$X^2 = \sum_{j=1}^k \frac{(N_j - n \cdot p_j)^2}{n \cdot p_j}$$

Si H_0 es cierta, se verifica que $X^2 \xrightarrow[n \rightarrow \infty]{D} \chi_{k-m-1}^2$, donde m es el número de parámetros estimados.

- e) Rechazar H_0 si $X^2 \geq \chi_{k-m-1, \alpha}^2$

Una consideración a hacer es que el valor del estadístico X^2 varía según los intervalos elegidos. No hay normas claras al respecto, excepto que no puede haber intervalos de frecuencia nula, pero, una idea es que sean de igual amplitud, que haya al menos 3 intervalos y que la frecuencia esperada sea al menos 5, es decir, $n \cdot p_j \geq 5$.

Test de Kolmogorov-Smirnov

Este test compara la función de distribución empírica de los datos con la función de distribución teórica.

Este test es aplicable en la versión que se presenta sólo para distribuciones continuas (existe una versión para distribuciones discretas que no se presenta en este capítulo), y es válido para cualquier tamaño de muestra, incluso aunque se disponga de pocos datos.

El procedimiento para aplicarlo es el siguiente:

a) Calcular la función de distribución empírica de los datos sin interpolación, es decir:

$$F_n(x) = \frac{\text{número de } X_i \text{'s } \leq x}{n}$$

b) Obtener la máxima diferencia entre la función de distribución empírica y la teórica: esta diferencia se alcanzará en los puntos observados, ya que son los puntos de salto de la función empírica, pudiendo ser en el propio punto o justo antes. Por ello, se calcula

$$D_n^+ = \max_{1 \leq i \leq n} \left| \frac{i}{n} - F(X_{(i)}) \right| \text{ que es la máxima diferencia en los puntos observados con el}$$

valor de la distribución empírica justo en el punto, y se calcula

$$D_n^- = \max_{1 \leq i \leq n} \left| \frac{i-1}{n} - F(X_{(i)}) \right| \text{ que es la máxima diferencia tomando el valor de la}$$

distribución empírica justo antes (por la izquierda) del punto. Así la máxima diferencia entre la función de distribución empírica y la teórica será el máximo de ambos valores,

$$\text{es decir, } D_n = \max \{ D_n^+, D_n^- \} .$$

c) Rechazar H_0 si $D_n > d_{n,1-\alpha}$, donde $d_{n,1-\alpha}$ es el valor que se recoge para esa significación y tamaño de muestra en las tablas de Kolmogorov-Smirnov.

En la siguiente figura se muestra gráficamente el significado de los valores D_n^+ y D_n^- .

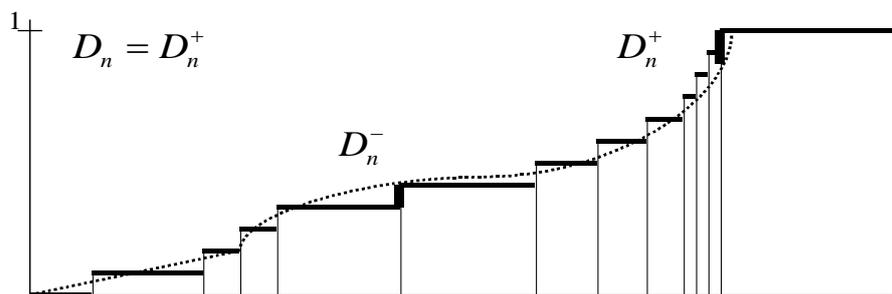


Figura 8. Representación gráfica de los valores D_n^+ y D_n^-

II.3 Generación de variables aleatorias

Una vez que se ha seleccionado una distribución para modelar la aleatoriedad de una variable de entrada, es necesario establecer procedimientos para obtener valores de esta variable durante la simulación del modelo. Así pues esta sección se dedica a la simulación de variables aleatorias con una distribución determinada.

Se empieza por generar valores de variables aleatorias con distribución uniforme en el intervalo $(0,1)$. La razón es que todos los métodos para generar variables aleatorias no son más que transformaciones de variables aleatorias con esta distribución y por ello el primer paso es saber obtener ésta.

II.3.1 Generación de muestras uniformes

Los números pseudoaleatorios son tan importantes que su generación no se puede dejar al azar.

Una secuencia de números se dice que es aleatoria si cualquier secuencia finita, seleccionada previamente a su diseño, es igualmente factible que esté incluida en aquélla.

Los mejores métodos para generar números aleatorios son los físicos y de entre ellos el mejor es el de la ruleta. Este método consiste en una ruleta dividida en 10 partes iguales, a las que se les asignan los valores del 0 al 9, y una flecha fija en un punto fuera de la ruleta. Si se la hace girar y posteriormente se la detiene bruscamente, se puede anotar el número que señala la flecha. Repitiendo esta operación n veces, se obtiene una secuencia de n números que obviamente constituye una secuencia de números aleatorios.

Estos valores se pueden considerar valores de una variable cuya distribución sea una uniforme discreta que toma los valores de 0 a 9 con probabilidad $1/10$ cada uno de ellos. Pero también los podemos agrupar de k en k y considerar que son valores de una uniforme discreta que toma valores $0, \dots, 10^k - 1$.

Si se desean generar valores de una uniforme en el intervalo $(0,1)$, podemos agrupar los valores de k en k y considerar que cada grupo son las cifras decimales de una realización de una variable aleatoria con distribución $U(0,1)$.

Obviamente, para considerar informativos los resultados obtenidos mediante la simulación de un modelo es necesario simularlo más de una vez. De hecho, se debe hacer una gran cantidad de veces, en general, más cuanto más complicado es el modelo, lo que hace ver la necesidad del uso del ordenador.

Existen tablas de números aleatorios obtenidos por el método de la ruleta y otros métodos físicos, pero no es un buen método para su uso en ordenador. Ésta fue la razón por la que se crearon métodos aritméticos particulares adaptados al ordenador, aunque con un cierto deterioro de la aleatoriedad, denominándose *números pseudoaleatorios*.

Un generador ideal de números pseudoaleatorios debe proporcionar secuencias de números con las siguientes propiedades:

- Tener distribución uniforme (son realizaciones de uniformes)
- Ser estadísticamente independientes
- Han de ser reproducibles
- Capaces de producir diferentes secuencias de números
- Deben tener un ciclo no repetitivo tan largo como se desee
- Ser generados rápidamente
- Ocupar poca memoria o almacenamiento en el ordenador

Veamos algunos métodos para generar números pseudoaleatorios.

Método de los cuadrados medios

Se toma un número al azar, x_0 , de $2n$ cifras. Se eleva al cuadrado y se toman de este resultado las $2n$ cifras centrales. Se repite el proceso.

Ejemplo:

$$x_0=4122 \quad x_0^2=16|9908|84$$

$$x_1=9908 \quad x_1^2=98|1684|64$$

$$x_2=1684 \quad x_2^2=2|8358|56$$

La secuencia 4122, 9908, 1684, ... puede ser considerada, al menos a partir de un cierto intervalo, como una secuencia de números pseudoaleatorios.

El principal problema de este método es que los números pueden repetirse a partir de una secuencia muy corta. Por ejemplo, si $x_0 = 3708$ se tiene que $x_4 = 6100 = x_8$, lo cual no es controlable.

Método de Lehmer

Sea x_0 un número al azar de n cifras. Se multiplica por otro k' (fijo del generador) de k cifras, dando lugar a uno de $n+k$. Se quitan las k cifras de la izquierda, obteniendo uno de n cifras al que se resta el de k cifras que se había separado.

Ejemplo:

$$x_0=4122 \quad k'=76 \quad 4122 \cdot 76=31|3272 \quad 3272-31=3241$$

$$x_1=3241 \quad k'=76 \quad 3241 \cdot 76=24|6316 \quad 6316-24=6292$$

Este método acaba degenerando a 0.

Métodos congruenciales

Estos métodos también son debidos a Lehmer (1951). Los métodos de generación de números pseudoaleatorios llamados congruenciales se basan en el concepto matemático de números congruentes.

Sean $x, y, m \in \mathbb{N}$. Se dice que x es congruente con y en módulo m si y sólo si x e y dan el mismo resto al dividir por m . Se representa por $x \equiv y \pmod{m}$.

Sean $a, b \in \mathbb{N} \cup \{0\}$, $m \in \mathbb{N}$. Se dice que una sucesión $\{x_n\}_{n \in \mathbb{N} \cup \{0\}}$ está generada por el método congruencial si y sólo si

$$x_{n+1} \equiv ax_n + b \pmod{m}$$

donde $x_n \in \{0, \dots, m-1\} \quad \forall n = 0, 1, \dots$, siendo x_0 la semilla de la sucesión que suele ser un valor dado por el programador, a m se le denomina *módulo* y a a *multiplicador*. Además si $b = 0$ se denomina generador multiplicativo y, en el caso general, generador mixto.

Los números generados por el método congruencial verifican que x_{n+1} es el resto de dividir $y_{n+1} = ax_n + b$ entre m . Además la ley recurrente es

$$x_{n+1} = y_{n+1} \bmod m = (ax_n + b) \bmod m \quad \forall n$$

Ejemplo: $m = 8$, $a = 5$, $b = 7$, $x_0 = 4$ genera la secuencia de números 4, 3, 6, 5, 0, 7, 2, 1, 4.

Ejemplo: $m = 9$, $a = 5$, $b = 1$, $x_0 = 1$

$y_1 = 5 \cdot 1 + 1 = 6$	$x_1 = 6$
$y_2 = 5 \cdot 6 + 1 = 31$	$x_2 = 4$
$y_3 = 5 \cdot 4 + 1 = 21$	$x_3 = 3$
$y_4 = 5 \cdot 3 + 1 = 16$	$x_4 = 7$
$y_5 = 5 \cdot 7 + 1 = 36$	$x_5 = 0$
$y_6 = 5 \cdot 0 + 1 = 1$	$x_6 = 1 = x_0$

Este método es evidente que genera valores a gran velocidad y ocupa poca memoria de ordenador, puesto que no se guarda toda la secuencia sino que para generar un valor sólo necesito el último valor generado. También es claro que es reproducible (con una misma semilla se obtiene una misma secuencia). Sin embargo, hay tres propiedades que no son tan evidentes y que no se cumplen para valores cualesquiera de a , b y m .

Respecto al ciclo no repetitivo en el mismo ejemplo se puede ver que es muy corto. De hecho, hay que observar que la máxima longitud que se puede lograr sin repetir es de m , ya que es la mayor cantidad de números diferentes que se puede obtener y una vez que se obtiene un número repetido toda la secuencia se repite, por lo que m hay que elegirlo suficientemente grande. Además se suele elegir de tal forma que facilite los cálculos (potencias de 2 o similares).

Existe una gran variedad de resultados buscando condiciones que han de tener los parámetros para que el ciclo no repetitivo sea el máximo posible.

A su vez, los generadores tienen que cumplir propiedades estadísticas: uniformidad e independencia de las secuencias. Para determinados valores de los parámetros de los generadores (valores para los que el ciclo sea el máximo posible) se han generado distintas secuencias y se les ha hecho pasar contrastes de bondad de ajuste para la uniformidad y contrastes de independencia.

De todo ello, se puede afirmar que los dos generadores multiplicativos siguientes generan valores con la propiedad de la uniformidad y la independencia y con ciclo no repetitivo maximal: $m = 2^{31} - 1$ para ambos y multiplicador $a = 16807$ o bien $a = 63036016$. El primero es más rápido y tiene menor riesgo de desbordamiento de memoria. Sin embargo, el segundo tiene mejores propiedades estadísticas⁵.

Por último, señalar que el fin de la generación de números pseudoaleatorios es obtener muestras uniformes en el intervalo (0,1), para ello lo que se hace es dividir cada número generado por el módulo m , con lo que la muestra obtenida puede ser considerada uniforme en ese intervalo⁶.

Si se dispone de diferentes cadenas de números aleatorios cada una se debe utilizar para un parámetro aleatorio [Law]. En este caso se debe tener en cuenta la longitud de la cadena para evitar solapes. En [Law] se pueden encontrar diferentes semillas para cadenas del generador mencionado separadas entre sí por 100000 números pseudoaleatorios. Algunas de estas semillas para el generador con parámetro $a = 63036016$ son las siguientes: 1973272912, 281629770, 20006270, 1280689831, 2096730329, 1933576050, 913566091.

Sin embargo, los generadores aleatorios anteriores con módulo alrededor de 2^{31} pueden ser insuficientemente largos para ciertas aplicaciones. De hecho, esta longitud se puede agotar en pocos minutos en un PC. Para suplir esta carencia se han diseñado otros generadores de mayor longitud y mejores propiedades, denominados *generadores múltiples recursivos combinados* [L'Ecuyer, 2002]. Estos generadores permiten la obtención simultánea de múltiples cadenas pudiendo cada una ser dividida en muchas subcadenas contiguas de gran longitud. Sus implantaciones en C, C++ o Java se pueden encontrar en la

⁵ En caso de ser necesaria la programación de un generador de números pseudoaleatorios hay que tener en cuenta que aunque el segundo de los valores se considera que tiene mejores propiedades estadísticas, computacionalmente da más problemas, sobre todo, de desbordamiento de memoria.

⁶ El valor uniforme en (0,1) se usará para generar otras variables aleatorias, pero para obtener el siguiente número de la secuencia ha de utilizarse el número pseudoaleatorio, es decir, el obtenido antes de dividir por m .

dirección www.iro.umontreal.ca/~lecuyer. El cálculo del número pseudoaleatorio u_n para una etapa n es la siguiente

$$\begin{aligned}x_{1,n} &= (a_1 x_{1,n-2} - b_1 x_{1,n-3}) \bmod m_1 \\x_{2,n} &= (a_2 x_{2,n-1} - b_2 x_{2,n-3}) \bmod m_2 \\z_n &= (x_{1,n} - x_{2,n}) \bmod 4294967087 \\u_n &= \begin{cases} z_n / 4294967088 & \text{si } z_n > 0 \\ 4294967087 / 4294967088 & \text{si } z_n = 0 \end{cases}\end{aligned}$$

siendo $a_1 = 1403580$, $b_1 = 810728$, $m_1 = 2^{32} - 209 = 4294967087$, $a_2 = 527612$, $b_2 = 1370589$ y $m_2 = 2^{32} - 22853 = 4294944443$. La longitud de este generador es $\rho = (m_1^3 - 1)(m_2^3 - 1) / 2 \approx 2^{191}$, muy superior a las anteriores.

Para generar diferentes cadenas y subcadenas se eligen dos enteros positivos v y w , se define $z = v + w$. Entonces, el ciclo ρ es dividido en cadenas contiguas de longitud $Z = 2^z$ y cada una a su vez es dividida en $V = 2^v$ subcadenas de longitud $W = 2^w$. Valores adecuados son $v = 51$ y $w = 76$, de manera que $W = 2^{76}$ y $Z = 2^{127}$. Para este generador en particular se pueden utilizar los siguientes valores como semillas iniciales por omisión

$$(x_{1,n-3}, x_{1,n-2}, x_{1,n-1}, x_{2,n-3}, x_{2,n-2}, x_{2,n-1}) = (12345, 12345, 12345, 12345, 12345, 12345).$$

Otros parámetros adecuados para generadores simulares se pueden encontrar en [L'Ecuyer, 1999].

Atención específica merece la generación de números aleatorios para ser utilizados en cálculos en paralelo [Fushimi].

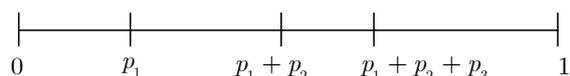
II.3.2 Generación de variables aleatorias discretas

II.3.2.1 Método general o estándar

Sea una variable aleatoria con una distribución discreta que toma ciertos valores con

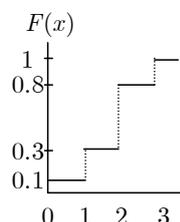
$$\text{determinada probabilidad } X = \begin{cases} x_1 & \text{con prob } p_1 \\ x_2 & \text{con prob } p_2 \\ x_3 & \text{con prob } p_3 \\ \vdots & \end{cases}, \text{ siendo } \sum_k p_k = 1.$$

La idea intuitiva del procedimiento es dividir el intervalo (0,1) en tantos subintervalos como valores puede tomar la variable y de tamaño las probabilidades de éstos. Generar un valor uniformemente distribuido en (0,1) y observar en qué subintervalo se encuentra y asignar a la variable el valor correspondiente a ese subintervalo.



Más formalmente, consiste en generar un número aleatorio uniformemente distribuido $u \in U(0,1)$, entonces $X = x_i$ si $\sum_{k=1}^{i-1} p_k \leq u < \sum_{k=1}^i p_k$, es decir, si $F_x(x_{i-1}) \leq u < F_x(x_i)$.

Por ejemplo, sea $X = \begin{cases} 0 & \text{con prob } p_1 = 0.1 \\ 1 & \text{con prob } p_2 = 0.2 \\ 2 & \text{con prob } p_3 = 0.5 \\ 3 & \text{con prob } p_4 = 0.2 \end{cases}$, con función de distribución



Para la secuencia de números aleatorios 0.27, 0.54, 0.06, 0.89 y 0.15, la variable tomará los valores siguientes: 1, 2, 0, 3 y 1.

Vamos a ver métodos concretos para algunas distribuciones, binomial y geométrica, sin olvidar que el método anterior sirve para cualquier distribución discreta. Para las demás distribuciones conocidas también hay métodos basados en su definición o en algunas propiedades, pero no se verán en este documento. Para ver una relación más extensa ver [Law] o [Bratley].

II.3.2.2 Binomial (n, p)

Se basa en la propiedad según la cual la distribución de la suma de v.a.i.i.d. con distribución de Bernoulli de parámetro p es Binomial (n, p). Luego, para generar valores de una variable con distribución Binomial (n, p), se pueden generar n Bernoulli de parámetro p y sumarlas.

Algoritmo:

1. $x \leftarrow 0$
2. Hacer n veces
Generar $u \in U(0,1)$
Si $u \leq p$ $x \leftarrow x+1$
3. Salida: X se distribuye según Binomial (n, p)

II.3.2.3 Geométrica (p)

La distribución geométrica corresponde al número de ensayo en que aparece el primer éxito al repetir un experimento de Bernoulli de parámetro p . Así que para generar valores con esa distribución es posible hacerlo con el método tradicional o aprovechando esta propiedad.

Con el método tradicional, basta aplicarlo con su función de distribución [$F(x) = 1 - (1-p)^x$ $x = 1, 2, \dots$] para ver que la expresión sería

$$x = 1 + \left[\frac{\ln(1-u)}{\ln(1-p)} \right] = 1 + \left[\frac{\ln(u)}{\ln(1-p)} \right]^\tau$$

Utilizando la propiedad que la relaciona con los experimentos de Bernoulli el algoritmo sería:

1. $x \leftarrow 0$
2. Hacer hasta que $u \leq p$
 $x \leftarrow x+1$

^{τ} La expresión obtenida es la primera, pero la utilizada es la segunda que tiene menos operaciones y la misma distribución, ya que si $u \stackrel{d}{=} U(0,1)$ entonces $1-u$ también.

Generar $u \in U(0,1)$

3. Salida: X se distribuye según Geométrica (p)

Este método es bueno cuando el valor de p es grande.

II.3.3 Generación de variables aleatorias absolutamente continuas

II.3.3.1 Método de la transformada inversa

Sea X la variable aleatoria cuya función de distribución es $F(x) = P\{X \leq x\}$. Se genera un número aleatorio uniforme entre 0 y 1, u , y luego se determina x tal que $F(x) = u$.

Supongamos que la variable tiene *distribución exponencial* con $F(x) = 1 - e^{-\alpha x}$ para $x \geq 0$ siendo $1/\alpha$ la media de la distribución. Dado un número aleatorio u tal que $F(x) = u$, luego $x = -\frac{\ln(1-u)}{\alpha} = -\frac{\ln(u)}{\alpha}$.

Otra aplicación directa de este procedimiento es para la *distribución uniforme* en un intervalo cualquiera (a,b) . La función de distribución en este caso es $F(x) = \frac{x-a}{b-a}$ si $x \in (a,b)$ (0 para valores menores y 1 para valores mayores) y dado un número aleatorio u tal que $F(x) = u$, se tiene que $x = a + (b-a)u$.

La *distribución Weibull* (α, β) es otra distribución para la que se puede aplicar este procedimiento directamente. La función de densidad de una distribución Weibull (α, β) de media $\frac{1}{\alpha\beta} \Gamma(1/\alpha)$, es $f(x) = \alpha\beta^\alpha x^{\alpha-1} e^{-(\beta x)^\alpha}$, $x \geq 0$, con lo que la función de distribución se obtiene de forma inmediata $F(x) = 1 - e^{-(\beta x)^\alpha}$, $x \geq 0$. Así, dado un valor aleatorio uniforme en $(0,1)$, u , el valor generado sería $x = \frac{1}{\beta} [-\ln(1-u)]^{1/\alpha} = \frac{1}{\beta} [-\ln(u)]^{1/\alpha}$.

Aunque éste es el procedimiento más extendido, sin embargo, muestra una dificultad fundamental para su aplicación, la necesidad de conocer explícitamente la función de distribución. La forma habitual de caracterizar una distribución absolutamente continua es mediante su función de densidad, de ahí que se hayan diseñado otros procedimientos basados en esta función.

II.3.3.2 Método de aceptación - rechazo

Se trata de un método general para variables absolutamente continuas. Existen dos versiones, la primera es más sencilla y con más limitaciones, pero también más utilizada precisamente por esa sencillez.

Método simple de rechazo

Se quieren obtener valores de una variable aleatoria con función de densidad $f(x)$ cuyo soporte es un intervalo acotado (a_1, a_2) . Sea c un valor tal que $c \geq \max \{f(x) : x \in (a_1, a_2)\}$.

La idea básica es generar puntos uniformemente en el rectángulo de base (a_1, a_2) y altura $(0, c)$. Si el punto está por encima de la curva el punto es rechazado y habrá que generar otro y si está por debajo se acepta su coordenada x como valor de una variable aleatoria con función de densidad $f(x)$.

El procedimiento es:

1. Generar $u_1, u_2 \in U(0,1)$

$$\text{Calcular } x = a_1 + (a_2 - a_1)u_1$$

$$\text{Calcular } y = cu_2$$

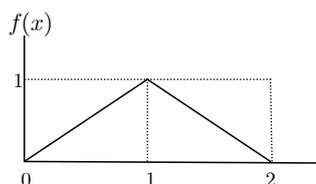
2. Calcular $f(x)$. Si $y > f(x)$ ir a 1
3. Salida: X toma el valor x que se distribuye según $f(x)$

Obsérvese que $P(\text{aceptar un valor dado por } (x, y)) = \frac{1}{c(a_2 - a_1)}$, por lo tanto el

valor de c es deseable que sea lo más pequeño posible. En particular, siempre que sea fácil de obtener se toma $c = \max \{f(x) : x \in (a_1, a_2)\}$.

Este procedimiento es especialmente relevante para distribuciones triangulares y trapezoidales. Veamos el siguiente ejemplo para una distribución triangular. Sea

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{fuera de } [0,2] \end{cases}$$



El procedimiento es el siguiente:

1. Generar $u_1 = U(0,1)$ y $u_2 = U(0,1)$

Calcular $x = 2u_1$ e $y = u_2$

2. Aceptar x si $y \leq f(x)$, rechazar si $y > f(x)$ y volver al paso 1

Este método tiene dos inconvenientes principales, que hacen que surja el método generalizado: que el soporte tiene que ser un intervalo acotado (hay muchas distribuciones que no cumplen esta hipótesis) y que la envolvente sea rectangular, cuando claramente puede haberlas mejores.

Método generalizado de rechazo

Sea $f(\cdot)$ una función de densidad con soporte no necesariamente finito de la que se desean obtener valores simulados. Sea $g(\cdot)$ una función de densidad elegida por nosotros tal que $\exists a > 1: f(x) \leq ag(x) \quad \forall x \in \mathbb{R}^s$ (en particular, esto implica que el soporte de $g(\cdot)$ ha de contener al soporte de $f(\cdot)$).

El procedimiento es el siguiente:

1. Generar $x = g(\cdot)$

^s $g(\cdot)$ se toma para que se puedan simular valores sin dificultad y previamente a aplicar el algoritmo es necesario calcular el valor de la constante a que haga que se verifique la relación anterior.

Generar $y \stackrel{d}{=} U(0, ag(x))$

2. Calcular $f(x)$ Si $y > f(x)$ ir a 1
3. Salida: X se distribuye según $f(x)$

Respecto a la elección de la envolvente $g(\cdot)$ no hay nada establecido acerca de cuál es la mejor, pero se debe elegir lo más parecida posible a la función a generar pero que sea sencillo obtener valores para ella. En cuanto al valor de la constante a , hay que tener en cuenta que $P(\text{aceptar } x) = P(y \leq f(x)) = 1/a$ y, por lo tanto, a debe ser lo menor posible pero manteniendo la hipótesis inicial, es decir, su valor óptimo sería $a = \sup \left\{ \frac{f(x)}{g(x)} : g(x) > 0 \right\}$.

Obsérvese también que el método simple visto anteriormente es un caso particular del método generalizado (con una distribución uniforme como envolvente).

Veamos un ejemplo: Supongamos que se desean generar valores de una distribución Normal(0,1) ($f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$), utilizando como envolvente la distribución logística

cuya función de densidad es $g(x) = \frac{e^{-x}}{(1+e^{-x})^2}$, y de aquí se deduce su función de

distribución de forma inmediata $G(x) = \frac{1}{1+e^{-x}}$, y mediante la transformada inversa para

generar valores de esta distribución basta calcular $x = -\ln \frac{1-u}{u}$, luego es una envolvente apropiada.

El primer paso, antes de empezar el algoritmo, es encontrar el valor de la constante

$$a = \sup \left\{ \frac{f(x)}{g(x)} : g(x) > 0 \right\} \text{ que resulta ser } a \approx 1.5958.$$

El algoritmo sería:

1. Generar $x \stackrel{d}{=} g(\cdot)$: Generar $u_1 \in U(0,1)$ y calcular $x = -\ln \frac{1-u_1}{u_1}$

Generar $y \stackrel{d}{=} U(0, ag(x))$: Generar $u_2 \in U(0,1)$ y calcular $y = 1.5958 g(x)u_2$

2. Calcular $f(x)$. Si $y > f(x)$ ir a 1
3. Salida: X se distribuye según $f(x)$

El principal inconveniente para utilizar este método es que no tiene una forma general ya que para cada distribución hay que elegir la envolvente más apropiada. Sin embargo, hay un caso en que es muy sencilla su aplicación, las distribuciones truncadas. En este caso, es claro que la mejor envolvente es la propia distribución sin truncar y que la aplicación del algoritmo se reduce a rechazar los valores obtenidos en la región que al truncar ya no forma parte del soporte y aceptar los demás.

Existen algunas distribuciones particulares para las que, por su relevancia, se han diseñado procedimientos específicos, en principio, más eficientes y generales que estos métodos.

II.3.3.3 Aplicación a una discreta: $Poisson(\lambda)$

La distribución de Poisson es una distribución discreta, sin embargo para generarla el mejor algoritmo se basa en el proceso estocástico de Poisson y su relación con la distribución exponencial, de ahí que no haya podido incluirse en la sección dedicada a las distribuciones discretas.

Sea $\{Y_t\}_{t \geq 0}$ un proceso estocástico. Se dice que $\{Y_t\}_{t \geq 0}$ es un proceso de Poisson de parámetro λ si y sólo si

i) $Y_0 = 0$ c.s.

ii) $\{Y_t\}_{t \geq 0}$ es un proceso de incrementos independientes, es decir, $\forall t, h > 0$ $Y_{t+h} - Y_t$ es independiente de $\{Y_s / 0 \leq s < t\}$

iii) $\{Y_t\}_{t \geq 0}$ es un proceso de incrementos estacionarios, cuya distribución es

$$\forall t, h > 0 \quad Y_{t+h} - Y_t \stackrel{d}{=} \wp(\lambda h)$$

Su aparición principal es en Teoría de Colas, donde se usa para contar el número de individuos que llegan a una cola hasta el instante t .

Propiedades del proceso de Poisson:

1. Un proceso de Poisson es un proceso de conteo

2. Sea $\forall t > 0$ Y_t : número de sucesos que ocurren en el intervalo $(0, t]$, y sea $\forall i = 1, 2, \dots$ X_i : tiempo entre los sucesos $i-1$ e i . $\{Y_t\}_{t \geq 0}$ es un proceso de Poisson de parámetro λ si y sólo si $X_i \stackrel{d}{=} \text{Exp}(\lambda) \quad \forall i = 1, 2, \dots$

Basándonos en esta última proposición, si $Y_t = k$ quiere decir que el número de sucesos ocurridos en $(0, t]$ es igual a k , o lo que es lo mismo, que $\sum_{i=1}^k X_i \leq t < \sum_{i=1}^{k+1} X_i$. Puesto que para $t=1$ la distribución de Y_1 es una Poisson de parámetro λ , se puede decir que un valor k generado de Y_1 es el primer valor natural k tal que $\sum_{i=1}^{k+1} X_i > 1$, donde X_1, X_2, \dots son v.a.i.i.d. con distribución $\text{Exp}(\lambda)$. De donde se deriva el siguiente algoritmo:

1. Hacer $x \leftarrow 0$, $Q = 0$
2. Generar $u = U(0,1)$. Hacer $Q = Q - \ln u$

Si $Q/\lambda > 1$ ir a 3). En otro caso, hacer $x \leftarrow x + 1$ y repetir el paso 2)

3. Salida: $X = \varphi(\lambda)$

II.3.3.4 Normal (μ, σ)

Para generar valores de una variable aleatoria con distribución $N(\mu, \sigma)$ basta con saber generar valores de la distribución $N(0,1)$, ya que es bien conocido que multiplicando ésta por σ y sumando μ se obtiene la distribución deseada. Por lo tanto, todos los esfuerzos se centran en saber obtener valores de la distribución $N(0,1)$.

Dado que no existe una expresión exacta para su función de distribución, el método de la transformada inversa no es aplicable en este caso. Aunque existe una aproximación funcional de la inversa de la función de distribución, vamos a ver ahora dos métodos particulares basados en resultados de la probabilidad.

Método del teorema central del límite

Este método se basa en el conocido teorema central del límite, por el cual si X_1, \dots, X_n son v.a.i.i.d. de media μ y desviación típica σ , entonces la distribución de $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ tiende a la distribución $N(0,1)$ cuando $n \rightarrow \infty$.

Este resultado aplicado a variables distribuidas según una distribución uniforme $U(0,1)$ sería $\frac{\sum_{i=1}^n u_i - n/2}{\sqrt{n/12}}$. Un valor que facilita los cálculos es $n=12$, con lo que el procedimiento sería obtener 12 variables con distribución uniforme $U(0,1)$ y calcular $\sum_{i=1}^{12} u_i - 6$.

El principal inconveniente del método es que el resultado en que se fundamenta es asintótico y, por lo tanto, se puede considerar válido un número grande, pero 12 no es precisamente un número grande.

Método de Box-Müller

Este método se basa en una transformación de variables aleatorias según la cual si u_1, u_2 son v.a.i.i.d., entonces $x = \sqrt{-2\ln u_1} \cos(2\pi u_2)$ e $y = \sqrt{-2\ln u_1} \sin(2\pi u_2)$ son v.a.i.i.d. con distribución $N(0,1)$. Así el procedimiento a seguir sería:

1. Generar $u_1, u_2 \in U(0,1)$
2. Salida $x = \sqrt{-2\ln u_1} \cos(2\pi u_2)$ e $y = \sqrt{-2\ln u_1} \sin(2\pi u_2)$ v.a.i.i.d. $N(0,1)$

Existe un método derivado del anterior, el *método polar de Marsaglia*, pero que no requiere la evaluación del seno o del coseno. En realidad, es un método de rechazo, ya que genera valores aleatorios en el cuadrado $(-1,1) \times (-1,1)$ y si éstos están dentro del círculo unidad expresa los valores anteriores en función de este punto y si no lo están los rechaza.

Algoritmo:

3. Generar $u_1, u_2 \in U(0,1)$

Calcular $v_1 = 2u_1 - 1$, $v_2 = 2u_2 - 1$, $w = v_1^2 + v_2^2$

4. Si $w > 1$ volver al paso 1

5. Hacer $y = \sqrt{\frac{-2 \ln w}{w}}$, $x_1 = v_1 y$, $x_2 = v_2 y$

6. Salida: x_1, x_2 v.a.i.i.d. según $N(0,1)$

Existen otros métodos para generar esta distribución, entre ellos existe una aproximación analítica de la inversa de la función de distribución, etc.

II.3.3.5 Lognormal (μ, σ)

La distribución lognormal, como su nombre indica, se deriva de la distribución normal, de modo que $Y = N(\mu, \sigma) \Leftrightarrow X = e^Y = LN(\mu, \sigma)$. El procedimiento para obtener valores de una variable con esta distribución consiste en generar la normal asociada y calcular la exponencial correspondiente.

Una información adicional sobre esta distribución que puede ser de interés en muchas ocasiones es el valor de esperanza y varianza de la distribución. Éstos son:

$$E[X] = e^{\mu + \sigma^2/2}$$

$$V[X] = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) = (e^{\sigma^2} - 1)E[X]^2$$

II.3.3.6 $\gamma(p, a)$ y Beta (α, β)

Vamos a estudiar estas variables conjuntamente, ya que están muy relacionadas. Para el caso de la Beta, se considerará la estándar, es decir, la que toma valores en el intervalo $(0,1)$; cualquier otra se obtiene multiplicando por la amplitud del intervalo final y trasladando. Separaremos en casos, según los parámetros sean naturales o no, ya que si los parámetros son naturales los procedimientos son bastante más sencillos y rápidos.

$\gamma(p, a)$ $p \in \mathbb{N}$

En este caso, se usa una propiedad por la cual si el parámetro p es natural $\gamma(p, a) = \text{Erlang}(p, a)$ y esta distribución Erlang se define como la suma de p exponenciales de parámetro a . Por lo tanto, el procedimiento será generar p v.a.i.i.d.

$$U(0,1), \text{ y calcular } x = -\frac{\sum_{i=1}^p \ln u_i}{a}.$$

Beta (α, β) $\alpha, \beta \in \mathbb{N}$

Se basa en que si $X \stackrel{d}{=} \gamma(\alpha, a)$ e $Y \stackrel{d}{=} \gamma(\beta, a)$ entonces $\frac{X}{X+Y} \stackrel{d}{=} \text{Beta}(\alpha, \beta)$, por lo que para generarla se obtienen $u_1, \dots, u_\alpha, v_1, \dots, v_\beta$ v.a.i.i.d. $U(0,1)$ y se calcula

$$B = \frac{\sum_{i=1}^{\alpha} \ln u_i}{\sum_{i=1}^{\alpha} \ln u_i + \sum_{i=1}^{\beta} \ln v_i}.$$

Caso general Beta (α, β) ($\alpha, \beta > 0$)

Se demuestra, mediante la pertinente transformación que el siguiente algoritmo genera valores de una distribución Beta:

1. Repetir hasta que $x + y \leq 1$
 Generar $u \in U(0,1)$. Hacer $x \leftarrow u^{1/\alpha}$
 Generar $v \in U(0,1)$. Hacer $y \leftarrow v^{1/\beta}$
2. Hacer $x \leftarrow \frac{x}{x+y}$
3. Salida: X se distribuye según $\text{Beta}(\alpha, \beta)$

Caso general $\gamma(p, a)$ ($p, a > 0$)

Igual que en el caso anterior, mediante transformaciones de variables aleatorias se demuestra que el siguiente algoritmo genera los valores deseados:

1. Hacer $n \leftarrow [p]$ y $r = p - [p]$
2. Generar $u_1, \dots, u_n \in U(0,1)$. Hacer $Z = -\sum_{i=1}^n \ln u_i$
3. Generar W $\text{Beta}(r, 1-r)$
4. Generar $u \in U(0,1)$. Hacer $Y = -\ln u$
5. Salida: $X = \frac{Z + WY}{a}$ se distribuye según $\gamma(p, a)$

II.3.3.7 χ_n^2

Para esta distribución se plantean dos métodos distintos, uno basado en la propia definición y otro en alguna de sus propiedades.

La definición de esta distribución es la siguiente: $X = \chi_n^2 \Leftrightarrow X = \sum_{i=1}^n X_i^2$, siendo

X_1, \dots, X_n v.a.i.i.d. $N(0,1)$.

Por lo tanto, un primer procedimiento es: Generar n variables $N(0,1)$, elevarlas al cuadrado y sumarlas. Sin embargo, este método puede ser muy lento.

La otra posibilidad es aprovechar la relación entre esta distribución y la Gamma: $\chi_n^2 = \gamma(p = n/2, a = 1/2)$

Distinguiremos dos casos según el parámetro p sea par o impar.

n par: $\chi_n^2 = \gamma(n/2, 1/2) = \text{Erlang}(n/2, 1/2)$. Algoritmo:

1. Generar $U_1, \dots, U_{n/2}$ v.a.i.i.d. $U(0,1)$

2. Salida: $X = -2 \sum_{i=1}^{n/2} \ln U_i = \chi_n^2$

n impar: $X = Y + X_n^2$ donde $Y = \gamma((n-1)/2, 1/2)$ y $X_n = N(0,1)$. Algoritmo:

1. Generar $U_1, \dots, U_{(n-1)/2}$ v.a.i.i.d. $U(0,1)$

Generar $Z \sim N(0,1)$

2. Salida: $X = -2 \sum_{i=1}^{(n-1)/2} \ln U_i + Z^2 = \chi_n^2$

II.3.3.8 t_n de Student

Basándonos en la definición, $T = t_n \Leftrightarrow T = \frac{Z}{\sqrt{Y/n}}$ donde $Z = N(0,1)$ e $Y = \chi_n^2$

independientes.

Hay que tener en cuenta que la T de Student como distribución para modelar variables aleatorias reales sólo se suele utilizar en finanzas, como una distribución semejante en forma a la normal de media 0 y desviación 1, pero cuyas colas son más pesadas. Sin embargo, para que

esto sea cierto y pueda sustituir a la normal, hay que dividir la t de Student por su desviación típica con el fin de que la variable resultante también tenga desviación típica la unidad. La varianza de una t de Student es $n/(n-2)$, por lo que la expresión resultante sería $\frac{T}{\sqrt{n/(n-2)}}$.

II.3.3.9 $F_{m,n}$ de Fisher-Snedecor

Mediante la definición: $F \stackrel{d}{=} F_{m,n} \Leftrightarrow F = \frac{\chi_m^2/m}{\chi_n^2/n}$ χ_m^2, χ_n^2 independientes.

II.3.4 Generación de variables aleatorias mixtas

Hasta ahora hemos visto cómo generar variables aleatorias cuya distribución es discreta o absolutamente continua. En esta sección vamos a ver un método general, que en particular, permitirá generar valores de variables cuya distribución no sea ninguno de los casos anteriores.

Teorema: Si U es una variable aleatoria $U(0,1)$ y $F(\cdot)$ una función de distribución arbitraria, la variable aleatoria definida por

$$Y = \inf \{z : U \leq F(z)\}$$

tiene por función de distribución $F(\cdot)$.

Algunos ejemplos de cómo se emplea este resultado son los siguientes:

- Variable absolutamente continua: el método aplicado a este caso coincide con el método de la transformada inversa.
- Variable discreta: en este caso es el mismo procedimiento que el del método general o estándar que vimos para distribuciones discretas, en el cual se dividía el intervalo en subintervalos de longitud las probabilidades.
- Variable mixta: sea la variable aleatoria mixta X cuya distribución viene definida por

$$F(x) = \begin{cases} 0 & x < 3/2 \\ x-1 & 3/2 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

La variable aleatoria $Y = \inf \{z : U \leq F(z)\}$ tendrá la misma distribución que X :

$$Y(u) = \begin{cases} 3/2 & 0 < u < 1/2 \\ u+1 & u \geq 1/2 \end{cases}$$

II.3.5 Generación de variables aleatorias multidimensionales

Las variables aleatorias multidimensionales si hay independencia entre las unidimensionales que la forman no requieren de métodos especiales. Por lo tanto, lo que resulta fundamental en este apartado en la relación de dependencia entre las variables, en muchos casos, como la normal, plasmada en la matriz de varianzas-covarianzas.

Por lo tanto, vamos a ver unas ligeras nociones de cálculo matricial para recordar algunos conceptos necesarios para entender los métodos que se presentarán posteriormente.

II.3.5.1 Nociones de cálculo matricial

Sea $A_{n \times n}$ una matriz cuadrada y $x_{n \times 1}$ un vector columna.

Definición: $x \neq 0$ es un *autovector* de A asociado al *autovalor* $\lambda \in \mathbb{C} \Leftrightarrow Ax = \lambda x$.

Observaciones:

1. $A_{n \times n}$ tiene exactamente n autovalores reales y complejos.
2. Cada autovalor tiene infinitos autovectores asociados
3. $Ax = \lambda x \Leftrightarrow Ax - \lambda x = 0 \Leftrightarrow (A - \lambda I)x = 0$, que tiene solución distinta de 0 si y sólo si $|A - \lambda I| = 0$.

Ejemplo:

$$A = \begin{pmatrix} 4 & 1 & -5 \\ -5 & 2 & 5 \\ 1 & 1 & -2 \end{pmatrix} \quad 0 = |A - \lambda I| = \begin{vmatrix} 4 - \lambda & 1 & -5 \\ -5 & 2 - \lambda & 5 \\ 1 & 1 & -2 - \lambda \end{vmatrix} = -\lambda^3 + 4\lambda^2 - \lambda - 6 = 0 \Leftrightarrow \\ \Leftrightarrow \lambda = -1, 2, 3$$

$$\text{Para } \lambda = -1 \quad (A - (-1)I)x = 0 \quad \begin{pmatrix} 5 & 1 & -5 \\ -5 & 3 & 5 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \\ \alpha \end{pmatrix} \quad \alpha \in \mathbb{R} \quad \dots$$

Propiedades:

1. Sea $A_{n \times n}$ simétrica. Entonces sus autovalores son reales y los autovectores correspondientes son ortogonales.

2. Diagonalización: Sea $A_{n \times n}$ simétrica. Entonces $A = CDC^{-1}$, siendo $D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$ la matriz diagonal de los autovalores y $C = [e_1, \dots, e_n]$ donde e_i : autovector columna asociado al autovalor λ_i de norma 1.

Definición: $A_{n \times n}$ es definida positiva $\Leftrightarrow x^T Ax > 0 \quad \forall x \neq 0$

Propiedades:

1. A simétrica definida positiva $\Leftrightarrow \lambda_1 > 0, \dots, \lambda_n > 0$
2. A simétrica definida positiva $\Rightarrow \exists P_{n \times n}$ tal que $A = PP^T$. Concretamente, $P = CB$ con

$$B = \begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \text{ y } C = [e_1, \dots, e_n] \text{ donde } e_i : \text{autovector columna asociado al autovalor}$$

λ_i de norma 1.

3. A definida positiva $\Rightarrow A^{-1}$ definida positiva
4. A definida positiva y B no singular $\Rightarrow B^T AB$ definida positiva
5. $A_{m \times n}$ no nula $\Rightarrow (A^T A)_{n \times n}$ definida positiva y simétrica

Triangulación de Cholesky: $\Sigma_{n \times n} = (\sigma_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, n}}$ simétrica y definida positiva \Rightarrow existe una

única matriz $A_{n \times n}$ triangular inferior tal que $\Sigma_{n \times n} = AA^T$. Algoritmo para obtener la matriz triangular de Cholesky:

1. Para $i = 1, \dots, n, j = 1, \dots, n \quad a_{ij} = 0$
2. Hacer $a_{11} = \sigma_{11}^{1/2}$ y para $i = 2, \dots, n \quad a_{i1} = \frac{\sigma_{i1}}{a_{11}}$
3. Para $j = 2, \dots, n$ hacer:

$$a_{ij} = \left(\sigma_{ij} - \sum_{k=1}^{j-1} a_{jk}^2 \right)^{1/2} \quad \text{Para } i = j+1, \dots, n \quad a_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} a_{ik} a_{jk}}{a_{jj}}$$

II.3.5.2 Normal Multivariante $N(\mu, \Sigma)$

Definición: $X = (X_1, \dots, X_n)^d = N(\mu, \Sigma) \Leftrightarrow f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$ donde

$\mu^T = (\mu_1, \dots, \mu_n)$ $\Sigma_{n \times n}$ matriz simétrica definida positiva

Corolario: $E[X] = \mu$ y Σ es la matriz de varianzas-covarianzas

Método 1 (Basado en la diagonalización)

Sabemos que $\Sigma = PP^T$ siendo $P = CB$ con $B = \begin{pmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}$ $C = [e_1, \dots, e_n]$ matriz

de autovectores normalizados.

Sean Z_1, \dots, Z_n v.a.i.i.d. $N(0,1)$, entonces $Z = (Z_1, \dots, Z_n)^d = N(0, I)$ siendo I la matriz identidad. En tal caso, se tiene que $X = \mu + PZ = N(\mu, \Sigma)$.

(El resultado es obvio ya que es normal y $E[X] = \mu + PE[Z] = \mu$ y $V[X] = E[(X - \mu)(X - \mu)^T] = PE[ZZ^T]P^T = PIP^T = PP^T = \Sigma$).

Algoritmo

1. Calcular $P = CB$
2. Generar Z_1, \dots, Z_n $N(0,1)$. Hacer $Z = (Z_1, \dots, Z_n)^T$
3. Salida: $X = \mu + PZ$

Método 2 (Basado en la triangulación de Cholesky)

Sea $Z = (Z_1, \dots, Z_n)^d = N(0, I)$ siendo I la matriz identidad, entonces $X = \mu + AZ = N(\mu, AA^T = \Sigma)$ donde A es la matriz triangular inferior de Cholesky

Algoritmo

1. Calcular A matriz triangular inferior de Cholesky
2. Generar $Z_1, \dots, Z_n \sim N(0,1)$. Hacer $Z = (Z_1, \dots, Z_n)^T$
3. Salida: $X = \mu + AZ$

Método 3 (Basado en las distribuciones condicionadas)

Se basa en que $f(x_1, \dots, x_n) = f(x_1)f(x_2/x_1)f(x_3/(x_1, x_2)) \cdots f(x_n/(x_1, \dots, x_{n-1}))$ y en el siguiente resultado:

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \stackrel{d}{=} N(0, \Sigma) \Rightarrow X_k / (X_1 = x_1, \dots, X_{k-1} = x_{k-1}) \stackrel{d}{=} N(a_k^T A_{k-1}^{-1} \omega_k, \sigma_k^2 - a_k^T A_{k-1}^{-1} a_k)$$

donde $A = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_k^2 \end{pmatrix}$, $a_k = \begin{pmatrix} \sigma_{1k} \\ \vdots \\ \sigma_{k-1,k} \end{pmatrix}$, $\omega_k = \begin{pmatrix} x_1 \\ \vdots \\ x_{k-1} \end{pmatrix}$

Algoritmo

1. Obtener las distribuciones de $X_1, X_2/X_1, X_3/(X_1, X_2) \cdots X_n/(X_1, \dots, X_{n-1})$
2. Para $k = 1, \dots, n$ generar x_k según $f(x_k/x_1 \cdots x_{k-1})$
3. Salida: $X = (x_1 + \mu_1, \dots, x_n + \mu_n) \sim N(\mu, \Sigma)$

De los 3 procedimientos el más aconsejable es el de la triangulación de Cholesky, pero los 3 son válidos, y en concreto el último es ampliamente utilizado.

Caso particular: Normal bivalente

Supóngase que se quiere simular (X_1, X_2) normal bivalente con $E[X_i] = \mu_i, i = 1, 2$, $V[X_i] = \sigma_i^2, i = 1, 2$ y $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$

Proposición: $(X_1, X_2) \stackrel{d}{=} N\left(\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right) \Rightarrow X_1 \stackrel{d}{=} N(\mu_1, \sigma_1^2)$ y

$$X_2 / (X_1 = x_1) \stackrel{d}{=} N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), \sigma_2^2 (1 - \rho^2)\right)$$

Algoritmo:

1. Generar $Z_1, Z_2 \sim N(0,1)$
2. Salida: $\left(X_1 = \mu_1 + \sigma_1 Z_1, X_2 = \mu_2 + \sigma_2 (\rho Z_1 + \sqrt{1 - \rho^2} Z_2)\right)$

II.3.5.3 Multinomial

Sea una población dividida en k clases con $P(A_i) = p_i, i = 1, \dots, k$. Se extraen n elementos con reemplazamiento, y se define X_i : número de elementos obtenidos de la clase $i, i = 1, \dots, k$. Se define la distribución multinomial como $(X_1, \dots, X_{k-1}) \stackrel{d}{=} M(n; p_1, \dots, p_{k-1})$

Método 1: basado en la definición

Algoritmo

1. Para $i = 1, \dots, k$ hacer $x_i = 0$
2. Repetir n veces

Generar J según una discreta tal que $P(Y = i) = p_i \quad i = 1, \dots, k$

Hacer $x_j \leftarrow x_j + 1$

3. Salida: (X_1, \dots, X_{k-1})

Método 2 o estándar

Las distribuciones condicionadas de una multinomial son: $X_1 \stackrel{d}{=} B(n, p_1),$

$$X_2 / X_1 = x_1 \stackrel{d}{=} B\left(n - x_1, \frac{p_2}{1 - p_1}\right), \dots, X_{k-1} / X_1 = x_1, \dots, X_{k-2} = x_{k-2} \stackrel{d}{=} B\left(n - \sum_{j=1}^{k-2} x_j, \frac{p_{k-1}}{1 - \sum_{j=1}^{k-2} p_j}\right)$$

Algoritmo

1. Hacer $x_0 = 0, p_0 = 0$

2. Para $i = 1, \dots, k-1$ generar $X_i \sim B(n - \sum_{j=0}^{i-1} x_j, \frac{p_i}{1 - \sum_{j=0}^{i-1} p_j})$

3. Salida: (X_1, \dots, X_{k-1})

II.3.5.4 T_n de Student multivariante

Para obtener muestras de una t_n t multidimensional de dimensión m con una matriz dada de correlaciones entre las componentes, Σ , podemos proceder como sigue.

Algoritmo

1. Generar (X_1, \dots, X_m) normal multivariante con vector de medias 0 y matriz de covarianzas Σ

2. Generar $Y = \chi_n^2$ independiente de las normales anteriores.

3. Para $j = 1, \dots, m$ hacer $t_j = \frac{X_j}{\sqrt{\frac{Y}{n}}}$

4. Salida: $(t_1, \dots, t_m) = T_n^d$ de Student multivariante de dimensión m y correlaciones Σ

III Software de simulación de modelos dinámicos

Es fácil darse cuenta de que a la hora de codificar modelos de simulación de eventos discretos hay algunas características que son comunes a todos los modelos, como es la generación de variables aleatorias con distribuciones específicas, la necesidad de implantar mecanismos de control y flujo del tiempo y la determinación del siguiente evento, la actualización de listas de sucesos (adición, supresión o actualización de registros), la recogida y análisis de datos generados por la simulación, la elaboración de informes, gráficas, etc.

La evidencia de esta alta intersección de características comunes es lo que llevó al desarrollo de lenguajes de simulación donde todas ellas estuvieran ya recogidas. Estos lenguajes están en una fase de gran desarrollo, al igual que toda la tecnología informática, lo que está propiciando el incremento del uso de la simulación como una herramienta de análisis de sistemas.

Sin embargo, como todo, el uso de los lenguajes de simulación tiene sus ventajas y sus inconvenientes.

III.1 Lenguajes de simulación versus lenguajes de propósito general

Desarrollar un modelo mediante un lenguaje de simulación obviamente presenta ciertas ventajas frente a desarrollarlo mediante un lenguaje de propósito general. Algunas de estas ventajas son las siguientes:

- Los lenguajes de simulación proporcionan de forma automática la mayoría de las características necesarias para programar un modelo de simulación, lo que redundará en una reducción de esfuerzo de programación y tiempo.
- Son un marco de trabajo natural para el uso de modelos de simulación, incorporando bloques básicos de programación más afines al desarrollo de un modelo.
- Son más sencillos de codificar y, por lo tanto, también son más fáciles de modificar para llevar a cabo los distintos experimentos.
- Facilitan la detección de errores, especialmente en la lógica del proceso de simulación.

Sin embargo, el desarrollo de un modelo con un lenguaje de propósito general presenta otro tipo de ventajas frente al uso de lenguajes de simulación:

- Son más generales y conocidos. En general, los modeladores conocen algún lenguaje de propósito general, pero no suelen conocer lenguajes de simulación.
- Son más accesibles, ya que estos lenguajes suelen estar disponibles en cualquier plataforma, no así los de simulación.
- Son más eficientes en requerimientos de memoria y suelen tener un tiempo de ejecución menor, ya que el lenguaje de simulación se diseña para gran variedad de problemas.
- Permiten mayor flexibilidad, ya que los lenguajes de simulación se diseñan con unas especificaciones que si bien intentan ser lo más generales posibles, siempre tienen limitaciones, mientras que los lenguajes de tipo general no las tienen.

Así pues, sabiendo las limitaciones de desarrollar un modelo de una u otra forma, la decisión ha de tomarse en las primeras fases del estudio, ya que condiciona tremendamente el modelo que se desarrolla. Entre otras cosas, si se opta por un lenguaje de simulación es importante conocer las posibilidades que éste tiene para poder establecer algunas hipótesis adicionales o modificar otras (por la falta de flexibilidad).

III.2 Tipos de software de simulación y características

Dentro del software de simulación, cabe distinguir dos tipos de software: los *lenguajes de simulación* y los *simuladores*.

Un *lenguaje de simulación* es un lenguaje de programación que es general por naturaleza pero con desarrollos especiales para cierto tipo de aplicaciones. Un modelo se desarrolla en un lenguaje de simulación escribiendo un programa mediante estructuras de modelado del lenguaje.

Un *simulador* es un programa que permite simular un sistema de una clase específica de sistemas con poca o ninguna programación. Para simular un sistema con un simulador se requiere poca o ninguna experiencia en programación, sin embargo, son muy limitadas las posibles configuraciones de los sistemas a simular. Ejemplos de simuladores son simuladores de vuelo, simulador de un centro de control de una central nuclear. El constructor del simulador ha de ser un programador, pero el usuario final no.

En todo software de simulación hay unas características que son deseables que éste contemple. Las principales características son:

- De tipo general:
 - Flexibilidad de modelado: con definición de atributos generales que sean fácilmente adaptables para distintos sistemas. En este sentido, un lenguaje de simulación es el software de simulación más flexible, mientras que los simuladores estarían en el extremo opuesto, aunque también entre ellos exista diferente grado de flexibilidad.
 - Facilidades para desarrollar y depurar el modelo incorporadas en el software. En particular las ayudas de depuración y la interactividad con el usuario son algunos de las características que más pueden facilitar el desarrollo del modelo.
 - Ejecución rápida del modelo, con una buena gestión de los eventos, interacción entre los distintos elementos, etc.
 - Admisión de modelos de gran tamaño, lo que supone una buena gestión de la memoria (memoria dinámica).
 - Disponibilidad en diferentes plataformas y portabilidad entre ellas (Windows, Linux, Unix, ...).
 - Posibilidad de combinar simulación discreta y continua, ya que es muy habitual que los modelos sean de tipo híbrido o combinado, con variables que cambian de forma continua y otras de forma discreta.
- Animación gráfica: La animación gráfica permite ver el sistema gráficamente con los cambios de estado. Aparte de que la presentación de esta forma sea mucho más atractiva, permite validar los modelos al ver gráficamente su funcionamiento, además de hacerlos más creíbles para el usuario final, ya que puede visualizar lo que está ocurriendo.
Dentro de la animación gráfica, hay dos tipos de animación: la animación en modo real o instantáneo (se presenta gráficamente al mismo tiempo que se hace la simulación) y la animación en modo “play-back” (se presenta la animación después de haber terminado la simulación). Este último modo se utiliza fundamentalmente porque la animación gráfica ralentiza enormemente la ejecución, así con este modo, primero se obtienen los resultados de forma eficiente y después se presenta gráficamente lo que ha ocurrido, pudiendo ser detenido en cualquier momento sin afectar a los resultados finales.
- Capacidades de estadística: el software debe tener características de estadística lo más amplias posible. Estas capacidades deben ser tanto para los datos de entrada (variedad en las distribuciones posibles, etc.) como para el análisis de los datos de salida (incorporar la posibilidad de hacer réplicas, obtener intervalos de confianza, incorporar técnicas de reducción de la varianza, etc.)
- Informes de resultados: debe admitir informes estándar o específicos, posibilidad de presentaciones tabulares o gráficos, permitir el acceso a muestras individuales, etc.)
- Soporte al cliente: debe ser extenso y bien documentado, incluyendo demostraciones y ensayos libres.

III.3 Enfoques o estrategias

La estrategia o visión del mundo de un lenguaje de simulación, inherente al lenguaje, es usada para seleccionar el evento siguiente y gestionar el tiempo. Básicamente, existen dos estrategias diferentes⁹:

- a) Programación de sucesos (Event Scheduling ES)
Algunos lenguajes de este tipo son GASP, SIMSCRIPT, SLAM o Arena.
- b) Interacción de procesos (Process Interaction PI)
De este tipo son GPSS, Q-GERT, SIMSCRIPT, SLAM, Arena o SIMULA.

Algunos admiten varias estrategias, pues su naturaleza y estructura permiten usar ambas.

Los programas realizados con un lenguaje de simulación tienen una estructura jerárquica con tres niveles:

1. **Nivel ejecutivo o del programa de control:** se encarga de identificar cuándo tiene que ocurrir el siguiente suceso y de la ejecución correcta de las operaciones por él implicadas en los momentos adecuados. El programa de control está implantado en el lenguaje propiamente dicho y el usuario del lenguaje no necesita saber cómo está programado, sólo necesita saber cómo funciona, es decir, qué estrategia tiene implantada.
2. **Nivel de operaciones:** es la secuencia de sentencias que constituyen el programa propiamente dicho, sentencias propias del lenguaje utilizado.
3. **Nivel de rutinas de detalle:** ejecutan las acciones implicadas por cada operación o sentencia del modelo y también son inherentes al propio lenguaje.

Con cualquiera de las estrategias, cuando se selecciona el suceso siguiente para ser procesado, se ejecuta la correspondiente rutina para modelar los cambios apropiados en el estado del modelo.

Entre los sucesos que cabe considerar, o mejor dicho, que un lenguaje de simulación puede considerar, se clasifican en dos tipos: sucesos incondicionales y sucesos condicionales.

⁹ Existe una tercera, programación de actividades, pero que está cayendo en desuso y por lo tanto no será mencionada.

- *Suceso incondicional*: es un suceso que es elegible para ser ejecutado cuando llega el instante de tiempo para el que ha sido programado. Depende exclusivamente del tiempo y de ninguna otra condición aparte de ésta.
- *Suceso condicional*: es un suceso que puede depender de condiciones adicionales distintas del tiempo para ser activado y que, por lo tanto, hay que verificar además del tiempo antes de ejecutar la correspondiente rutina; esas condiciones suelen ser condiciones relativas al estado de las componentes del sistema (desocupación o fin de ocupación de un dispositivo del sistema, ...).

Vamos a ver ahora en qué consiste cada una de las estrategias de los lenguajes de simulación.

1. PROGRAMACIÓN DE SUCESOS (*Event Scheduling*, ES)

Esta estrategia considera que es una secuencia de sucesos incondicionales a lo largo del tiempo, es decir, no se programa ningún suceso que dependa de condiciones adicionales al tiempo. Se selecciona de la lista de sucesos aquél cuyo tiempo de ocurrencia es el más próximo, resolviendo los empates por las prioridades asignadas o por omisión, se actualiza el reloj de simulación igualando su valor al del instante en que ocurre el suceso y se llama a la rutina correspondiente al tratamiento del suceso. La traza del modelo sigue esta estrategia de programación de sucesos.

Cualquier verificación de una condición diferente de la del tiempo del reloj debe realizarse dentro de las rutinas de tratamiento de los sucesos y durante la ejecución de las rutinas de evento no se avanza el reloj de simulación. Por ejemplo, mediante la asignación de valores singulares a ciertas variables auxiliares, tal como se hacía asignando valor ∞ al instante del próximo evento de final de servicio cuando no había clientes en el sistema.

Esta filosofía responde al planteamiento que se ha hecho previamente. Se podría decir que el programa principal y la rutina de tiempo es lo que corresponde al nivel ejecutivo en este caso y las rutinas de evento sería lo único que el modelador tendría que programar.

Por ejemplo, en un sistema de colas con un servidor, los eventos incondicionales son llegadas de clientes y fin de servicio de clientes y el programa en un lenguaje de simulación serían las rutinas de evento de cada uno de ellos.

2. INTERACCIÓN DE PROCESOS (*Process Interaction*, PI)

Contempla el progreso de las componentes del sistema a través de una secuencia de pasos o procesos, cada uno de los cuales puede tener dos posibles componentes, un segmento de condición cuya ejecución identifica si se puede pasar a ejecutar la segunda componente, que es un segmento de acción.

Un proceso es una secuencia ordenada de eventos interrelacionados, separados por “pasos” de tiempo, que describe la experiencia entera de una entidad como su flujo a través de un sistema. Así, un proceso, o mejor dicho, una rutina de proceso contiene explícitamente el camino del tiempo simulado de cada entidad y suele tener múltiples puntos de entrada. Un modelo de simulación puede tener diferentes tipos de procesos.

Para el ejemplo de un sistema de colas con un servidor, el proceso comienza por la llegada de un cliente (que sería una entidad), cómo este ocupa al servidor o es situado en espera, la ocupación del servidor y tiempo que lo tiene ocupado y el final de servicio y desaparición de la entidad del sistema.

Las *entidades* del sistema pueden ser fijas o móviles. Las entidades fijas (o *recursos*) son denominadas procesadores y representan servidores en un sistema de colas, instalaciones, semáforos y entidades lógicas. Las entidades móviles (o *cargas*), o simplemente entidades, son las que se mueven a través del proceso.

En un instante determinado cada entidad se puede encontrar en un estado diferente. Los posibles estados de una entidad son los siguientes:

- *Activa*: entidad que se mueve a través de su ruta o proceso en ese momento.
- *Demorada*: entidad que se encuentra retenida por un suceso *incondicional* y, por lo tanto, tiene previsto un tiempo de activación.
- *Detenida*: se encuentra retenida por un suceso *condicional*, no tiene instante previsto de activación, éste depende del cambio de estado de alguna componente.

En el ejemplo de un sistema de colas con un servidor, como ya se ha dicho, las entidades son los clientes. Cuando se está ejecutando el proceso de un cliente y avanzando éste por la rutina del proceso se dice que esa entidad está activa (sólo una está activa en un instante determinado). Cuando el cliente está siendo atendido por el servidor, la entidad se considera que está demorada y cuando el cliente está esperando en la cola a que el servidor se libere y él sea el primero en ser atendido, se considera que es una entidad detenida. Gráficamente, el proceso de un sistema como el descrito tendría que considerar los siguientes puntos:

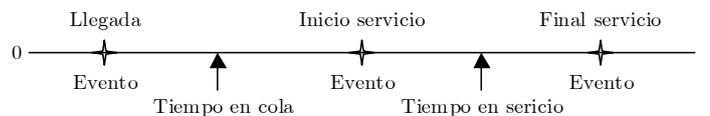


Figura 9. Proceso de un cliente en una cola con un servidor.

Cuando una entidad está activa se mueve por su ruta o proceso hasta que es detenida o demorada por algún suceso o desaparece del sistema. En ese momento se vuelve al programa de control que analiza las entidades que hay abiertas en ese momento (u otras que puedan surgir) y, ante los cambios producidos por el proceso de la entidad que acaba de dejar de ser activa, cuál de todas las entidades es la primera que se puede activar, moviendo el reloj de simulación al instante en que se activa esta entidad.

Un diagrama del proceso de un cliente en una cola con un servidor puede verse en la Figura 10. Diagrama del proceso en una cola con un servidor. Obsérvese que el punto 4 supone una espera condicional y el punto 8 una espera incondicional. Así mismo, los puntos 1, 5 y 9 son puntos de entrada desde el programa de control.

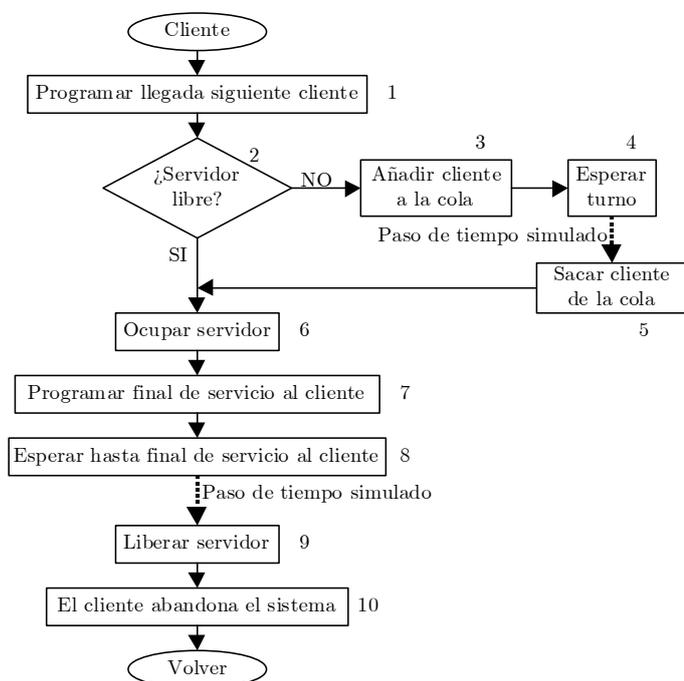


Figura 10. Diagrama del proceso en una cola con un servidor

IV Análisis de resultados

Un estudio de simulación busca respuestas a preguntas sobre el sistema objeto del estudio a través de la información que proporcionan los experimentos con el modelo del sistema. A su vez los experimentos buscan, en general, respuestas a preguntas del tipo: ¿Qué pasaría sí? (*What-if*) que se plantean en distintas fases del ciclo de vida: diseño, modificaciones de sistemas ya existentes, ...

Las respuestas que buscamos mediante los experimentos servirán de soporte para tomar una decisión racional sobre el sistema. Así pues, interesa que las respuestas queden expresadas numéricamente para la alternativa que nos ocupa en cada momento.

Las alternativas constituirán una variante del modelo o escenario de simulación con las que realizaremos los experimentos, y con la que obtendremos una estimación de las variables respuesta.

Como tales estimaciones, siempre que el modelo incluya aleatoriedad habrá que recurrir para analizarlas y obtener conclusiones a la estadística, y más concretamente a los métodos de muestreo, los métodos de reducción de la varianza, los métodos de estimación y al diseño de experimentos.

IV.1 Comportamientos transitorio y estacionario de un proceso estocástico

En los modelos de simulación dinámicos y aleatorios la variable respuesta varía con el tiempo, por lo tanto puede ser realmente considerado un proceso estocástico.

Así pues comenzaremos explicando brevemente los dos tipos de comportamiento que puede tener un proceso estocástico: el transitorio y el estacionario o permanente.

Considérese el proceso estocástico respuesta Y_1, Y_2, \dots , y sea $F_i(y/I) = P(Y_i \leq y/I)$, $y \in \mathbb{R}$ la distribución de la variable i -ésima del proceso supuestas unas condiciones iniciales I . $F_i(y/I)$ es denominada *distribución transitoria del proceso en el instante i para las condiciones iniciales I* .

Si cuando el tiempo se hace tender a ∞ , la distribución deja de depender del tiempo y de las condiciones iniciales, es decir, $F_i(y/I) \xrightarrow{i \rightarrow \infty} F(y) \quad \forall y, \forall I$, entonces se dice que $F(y)$ es la *distribución estacionaria*.

Es un límite, pero en la práctica suele existir un cierto k a partir del cual las distribuciones son casi iguales.

Las siguientes gráficas muestran la evolución de un cierto proceso estocástico que alcanza un estado estacionario. En la primera, se representa la distribución de la variable respuesta y como va cambiando con el tiempo. En la segunda, se muestra la evolución de la media de la variable respuesta (número de clientes en un sistema de colas con un servidor) con distintas condiciones iniciales.

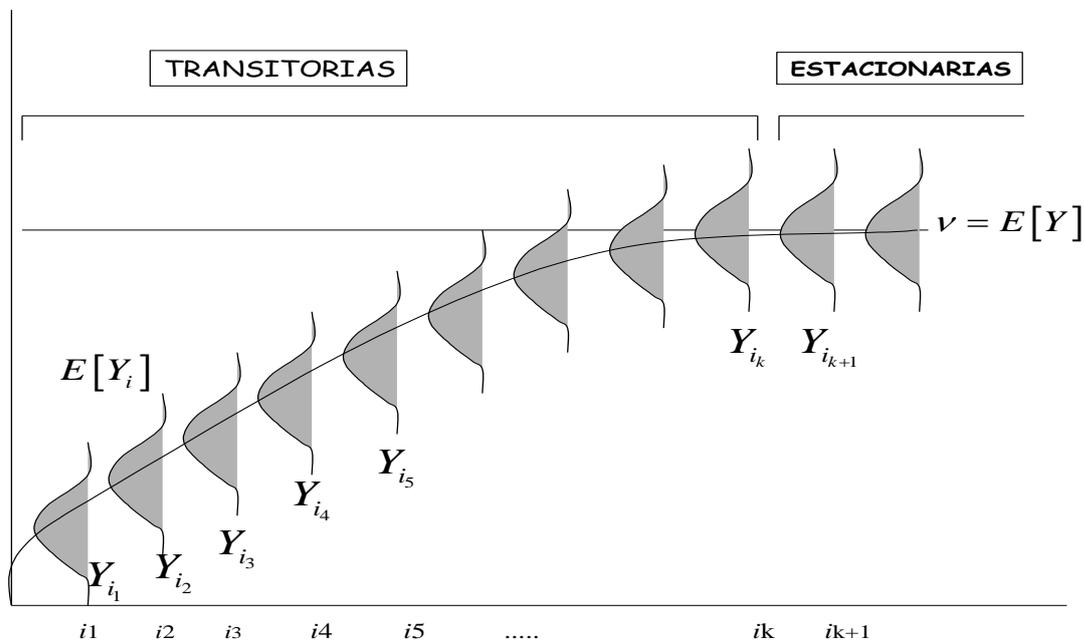


Figura 11. Evolución de la distribución de la variable respuesta

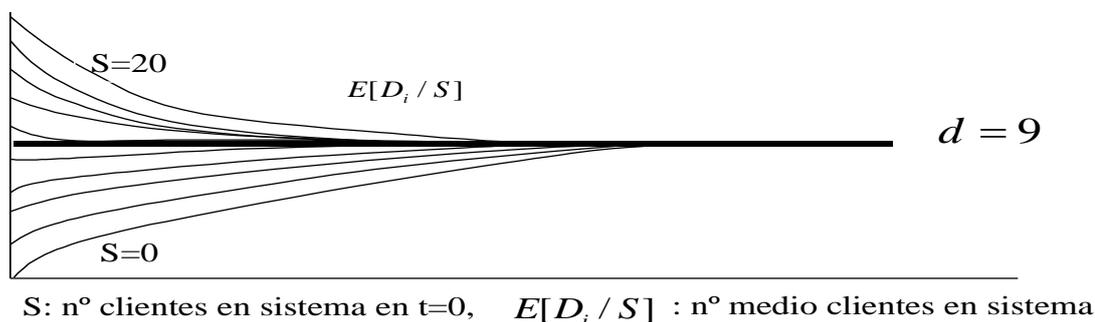


Figura 12. Evolución de la media de la variable variando condiciones iniciales

IV.2 Tipos de simulación según el análisis de resultados

Cabe distinguir varios tipos de simulación según sea el tipo de análisis que se pretenda hacer, o más concretamente, como sea el proceso estocástico de la variable respuesta y qué características de su comportamiento se pretenden analizar según su horizonte temporal. Así los tipos de simulación que se pueden plantear son:

- *Simulación con horizonte finito*: es la que se lleva a cabo cuando existe un evento “natural” E que especifica la longitud de cada simulación o replicación. En ese evento el sistema se reinicializa, obteniendo una muestra aleatoria simple de variables respuesta. Las condiciones iniciales generalmente afectan a las medidas de desarrollo, por lo que han de ser representativas del sistema real, para lo cuál se puede plantear un periodo de “calentamiento” o “arranque” (warm up) en el que se alcancen esas condiciones ó se puede modelar las condiciones iniciales y aleatorizar en cada replicación.

En ocasiones, cuando el horizonte es muy lejano respecto al periodo de arranque, siendo el comportamiento estacionario el que rige la mayor parte del tiempo, la simulación con horizonte finito se puede asimilar a una simulación con horizonte infinito.

- *Simulación con horizonte infinito*: no existe tal evento que indique el final de la replicación, y nuestro interés se centra en el comportamiento a largo plazo, pudiendo entonces darse varias posibilidades:
 1. *Existe distribución estacionaria*: el objetivo es estimar los parámetros estacionarios del modelo.
 2. *No existe distribución estacionaria, pero sí por ciclos*: entonces hay que estimar los parámetros estacionarios de cada ciclo y la duración de éstos.

3. *No existe distribución estacionaria, pues los datos de entrada varían en el tiempo:* entonces hay que considerar que cada vez que cambian es un final de horizonte, y tratarlo así, como si fueran sucesiones de simulaciones con horizonte finito.

IV.3 Estimación de variables respuesta: precisión y tamaño muestral

En general, el valor esperado de la variable respuesta se estima mediante la media muestral de las observaciones, que es el estimador puntual para ésta. Sin embargo, en muchas ocasiones no es suficiente dar un valor puntual, sino que se desea también dar un intervalo de confianza para ésta (estimación por intervalo) o una precisión del valor obtenido.

Así, dada una muestra aleatoria simple de la variable respuesta, Y_1, \dots, Y_n , el estimador

puntual para la media es la media muestral, $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$, que es un estimador insesgado o centrado para la esperanza, es decir, $E[\bar{Y}] = E[Y]$.

Por otra parte, un estimador para la varianza de la variable Y es la cuasivarianza, cuya

expresión es $S^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}$, de modo que dado que la varianza de la media muestral es

$Var[\bar{Y}] = \frac{\sigma^2}{n}$, entonces, el valor $\frac{S^2}{n}$ resulta ser un estimador para la varianza de la media muestral. Estos estimadores serán insesgados sólo si la muestra es aleatoria simple, es decir, si existe independencia entre las variables.

Los valores de la media y varianza muestrales se pueden calcular iterativamente de acuerdo con estas expresiones [Ramos:90]

$$\bar{Y}_n = \frac{1}{n} \left[(n-1)\bar{Y}_{n-1} + Y_n \right]$$

$$S^2(n) = \frac{1}{n-1} \left[(n-2)S^2_{(n-1)} + \frac{n}{n-1} (\bar{Y}_n - Y_n)^2 \right]$$

$$S^2(n) = \frac{1}{n-1} \left[(n-2)S^2_{(n-1)} + \frac{n-1}{n} (\bar{Y}_{n-1} - Y_n)^2 \right]$$

o estas expresiones numéricamente estables

$$S^2(n) = \frac{1}{n-1} \left[(n-2)S^2_{(n-1)} + \frac{1}{n(n-1)} \left(\sum_{i=1}^{n-1} Y_i - (n-1)Y_n \right)^2 \right]$$

$$S^2(n) = \frac{1}{n-1} \left[(n-2)S^2_{(n-1)} + \frac{1}{n(n-1)} \left(\sum_{i=1}^n Y_i - nY_n \right)^2 \right]$$

siendo Y_n el valor de la última observación y tomando como valores iniciales $\bar{Y}_1 = Y_1$ y $S^2(1) = 0$.

Con estas consideraciones, dado un nivel de confianza $1 - \alpha$, el intervalo de confianza para la esperanza de la variable respuesta, suponiendo distribución normal o una muestra suficientemente grande, tiene la siguiente expresión:

$$\bar{Y} \pm t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}$$

donde $t_{n-1, 1-\alpha/2}$ es el valor de la t de Student con $n-1$ grados de libertad que deja una probabilidad de $\alpha/2$ a su derecha.

Su interpretación es que da cada 100 intervalos que construyéramos, confiamos en que en al menos en $(1-\alpha) \cdot 100$ se encontrará la esperanza de la variable. Obsérvese que al hacer sólo uno, no hay seguridad de que la media real se encuentre en tal intervalo, sólo cierta confianza.

Por otra parte, si se desea dar una estimación de la precisión de la media, se puede deducir a partir del intervalo de confianza. A la vista de la expresión anterior una multiplicación por 4 del número de muestras implica una disminución del intervalo de confianza aproximadamente a la mitad.

Por lo tanto, una medida de la *precisión de la estimación de la media* viene dada por la relación entre la desviación típica de la media y la media

$$\beta = \frac{\sqrt{\frac{S^2(n)}{n}}}{\bar{X}_n}$$

A partir de aquí, se pueden plantear dos tipos de muestreo:

- *Muestreo de dimensión fija*: el tamaño de la muestra, n , se fija de antemano, y se obtiene una precisión no controlada previamente.
- *Muestreo secuencial*: la precisión es fijada de antemano y lo que queda indeterminado y sujeto a los resultados que se vayan obteniendo es el tamaño de la muestra; en cada iteración se calcula la precisión lograda, y si no alcanza a la prefijada se sigue muestreando.

IV.4 Estimación de parámetros estacionarios: el problema del estado inicial transitorio

Como ya se ha comentado en un proceso estocástico, como es la variable respuesta de un modelo de simulación dinámico, existe un comportamiento transitorio, y si el proceso es estacionario, otro comportamiento que es el que se alcanza en el periodo estacionario.

Si la simulación es de *horizonte finito*, el periodo transitorio ha de tenerse en cuenta, de tal forma que incluso el objetivo del análisis sea estudiar el comportamiento en este periodo.

Sin embargo, si la simulación es de horizonte infinito, o se pretende estimar un parámetro estacionario o de comportamiento “normal”, es decir, se quiere estimar $\nu = \lim_{i \rightarrow \infty} E[Y_i]$, hay que evitar las influencias del estado inicial.

Algunos métodos para obtener una muestra para la media estacionaria y un estimador puntual para ésta son los siguientes:

- A) Replicación/Eliminación
- B) Procedimiento por Lotes
- C) Procedimientos regenerativos

A) Método de Replicación/Eliminación:

Se realizan n replicas independientes de longitud m . Se determina el periodo de arranque (warmup) de longitud l ($l \ll m$), y se eliminan esas observaciones en la estimación. Así si las observaciones de cada replicación son $Y_{11}, \dots, Y_{1m}, \dots, Y_{n1}, \dots, Y_{nm}$, se consideran sólo las observaciones posteriores al periodo de arranque para hacer la estimación, obteniéndose de

cada replicación $X_1 = \frac{\sum_{i=l+1}^m Y_{li}}{m-l}$, ..., $X_n = \frac{\sum_{i=l+1}^m Y_{ni}}{m-l}$, y posteriormente haciendo la media y su intervalo con estos valores: $\bar{X} \pm t_{n-1, \alpha/2} \frac{S_x}{\sqrt{n}}$

El problema de este método es que puede existir una desviación entre el valor estimado, $\hat{\nu}$, y el verdadero valor de ν .

En cuanto a la aplicación presenta la dificultad de elegir la longitud de periodo de arranque, lo que se suele hacer mediante un análisis previo de tipo gráfico mediante medias móviles. Por otra parte, dependiendo de cuál sea esa longitud, resulta poco eficiente simular el comportamiento desde el principio cada vez, para luego prescindir de las primeras observaciones.

B) Procedimiento por Lotes:

En este caso se realiza una única replicación, con lo que sólo hay un periodo de arranque, que debe ser eliminado. El resto de la secuencia se divide en n lotes de tamaño k . k y n deben ser “suficientemente” grandes, de modo que los bloques se puedan considerar independientes y sus medias distribuidas normalmente (para que sea aplicable el Teorema Central del Límite para obtener el estimador.

Así, suponiendo que las observaciones, una vez eliminado el periodo de arranque, fueran $Y_1, \dots, Y_k, \dots, Y_{nk}$, se dividen en los lotes de los que se obtiene la media:

$$\underbrace{Y_1, \dots, Y_k}_{\bar{Y}_1(k)}, \underbrace{Y_{k+1}, \dots, Y_{2k}}_{\bar{Y}_2(k)}, \dots, \underbrace{Y_{(n-1)k+1}, \dots, Y_{nk}}_{\bar{Y}_n(k)}.$$

Por último, se obtiene la media de los valores medios obtenidos de cada lote

$$\bar{Y}(n, k) = \frac{\sum_{j=1}^n \bar{Y}_j(k)}{n} = \frac{\sum_{i=1}^{nk} Y_i}{nk}$$

y la cuasivarianza para poder hacer el intervalo de confianza o la precisión

$$S^2(n, k) = \frac{\sum_{j=1}^n (\bar{Y}_j(k) - \bar{Y}(n, k))^2}{n-1}$$

El problema que puede presentar este método es que se haga una estimación baja de la varianza del estimador, $Var(\hat{V})$. Esta infraestimación puede darse porque al ser una única replicación las observaciones no son independientes y, por lo tanto, si el tamaño de cada replicación no es suficientemente grande puede haber correlación entre ellas y no ser correcta la estimación de la varianza. Ésa es, por lo tanto, la mayor dificultad del método, cómo obtener el tamaño k para lograr la incorrelación. Respecto al método anterior presenta la ventaja de no tener que realizar la simulación del periodo de arranque más que una vez.

C) *Procedimientos regenerativos:*

En este caso se realiza una única replicación, como antes, pero los bloques no son del mismo tamaño, sino que se toma un punto de regeneración para determinar dónde acaba, siendo variable la longitud de cada secuencia obtenida, N_j . Se entiende por *punto de regeneración*, un punto en el que el sistema vuelve a comenzar como si fuera el principio. Por ejemplo, en un sistema de colas podría ser que el sistema se encuentre vacío. De ese modo, se evita el problema de la correlación entre las observaciones de dos secuencias. Sin embargo, ahora hay que tener en cuenta que el tamaño de cada secuencia es también una variable aleatoria.

Así si las observaciones obtenidas, agrupadas ya en secuencias, son las siguientes

$$Y_1, Y_2, \dots, Y_{B_1}, Y_{B_1+1}, \dots, Y_{B_2}, \dots, Y_{B_{n-1}+1}, \dots, Y_{B_n}$$

con tamaños respectivos, $N_1 = B_1$, $N_2 = B_2 - B_1$, $N_n = B_n - B_{n-1}$, se consideran para cada

secuencia dos variables: la suma de las observaciones, $X_j = \sum_{i=B_{j-1}+1}^{B_j} Y_i$, y el tamaño de éstas, N_j .

Para cada una de estas variables se consideran las estimaciones de su media y varianza, así a partir de las X_j se obtienen \bar{X} y S_X^2 , y a partir de las N_j se obtienen \bar{N} y S_N^2 .

Por último, el estimador de la media será $S^2 = S_X^2 - 2\bar{Z}S_{X,N}^2 + \bar{Z}^2S_N^2$, el estimador de su

varianza, será $S^2 = S_X^2 - 2\bar{Z}S_{X,N}^2 + \bar{Z}^2S_N^2$, donde $S_{X,N}^2 = \frac{\sum_{j=1}^n (X_j - \bar{X})(N_j - \bar{N})}{n-1}$, y el intervalo

de confianza resultante $\bar{Z} \pm z_{\alpha/2} \frac{S}{\bar{N}\sqrt{n}}$.

Con este método también se corre el riesgo de hacer una infraestimación de la varianza. Pero su principal inconveniente es que puede ser difícil de aplicar, bien porque no haya punto de regeneración o que la probabilidad de que se dé sea muy pequeña teniendo que hacer largas simulaciones para lograr una muestra o que el número de secuencias obtenido sea muy pequeño, o bien por todo lo contrario, que resulten secuencias de tamaño N_j demasiado pequeños, que hagan que no se puedan considerar independientes las secuencias y por lo tanto se esté infravalorando su varianza.

IV.5 Estimación de integrales por Monte Carlo

El objeto de un estudio de simulación suele ser la estimación de una o más esperanzas de la forma

$$\theta = E[\Phi(X)] = E[\Phi(X_1, \dots, X_k)]$$

donde $\Phi: (\mathbb{R}^k, \mathcal{B}^k) \rightarrow (\mathbb{R}, \mathcal{B})$ medible y $X = (X_1, \dots, X_k)$ es una variable aleatoria k -dimensional.

No es sorprendente entonces que se use la simulación para calcular integrales múltiples sin hacer mención de ningún modelo estocástico. Esto se denomina *integración de Monte Carlo*.

Sin embargo, el modelo original puede ser estocástico o no en sí mismo, y las técnicas que se van a mostrar son válidas en ambos casos.

IV.5.1 Método Monte Carlo de acertar o fallar

Supongamos que se quiere estimar $\theta = \int_a^b \Phi(x) dx$ $\Phi: (a, b) \rightarrow \mathbb{R}^+$.

Sea c una constante tal que $c \geq \Phi(x) \quad \forall x \in (a, b)$.

Se simulan n valores uniformes en el rectángulo $[a, b] \times [0, c]$, y sea

R : número de puntos simulados que caen debajo de la curva $\stackrel{d}{=} B(n, \frac{\theta}{c(b-a)})$

El estimador $\tilde{\theta} = c(b-a) \frac{R}{n}$ verifica que $E[\tilde{\theta}] = \theta$ y $V[\tilde{\theta}] = \frac{\theta}{n} [c(b-a) - \theta]$

IV.5.2 Método de Monte Carlo crudo

Sea $\Phi : (\mathbb{R}^k, \mathcal{B}^k) \rightarrow (\mathbb{R}, \mathcal{B})$ medible.

Supongamos que se quiere estimar $\theta = E[\Phi(x)] = \int_{\mathbb{R}^k} \Phi(\vec{x}) f(\vec{x}) d\vec{x}$.

El estimador de Monte Carlo se define como

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \Phi(X^{(i)})$$

donde $X^{(1)}, \dots, X^{(n)}$ es una muestra aleatoria simple simulada de $X = f(\vec{x})$

Propiedades:

1. $E[\hat{\theta}] = \theta$ $V[\hat{\theta}] = \frac{1}{n} \int_{\mathbb{R}^k} (\Phi(\vec{x}) - \theta)^2 f(\vec{x}) d\vec{x} = \frac{1}{n} V[\Phi(X^{(i)})]$
2. $\Phi : (a, b) \rightarrow \mathbb{R}^+$, $a, b \in \mathbb{R}$, $0 \leq \Phi \leq c$, $\theta = \int_a^b \Phi(x) dx \Rightarrow V[\hat{\theta}] \leq V[\tilde{\theta}]$, dándose la igualdad sólo si la función es una constante en el intervalo (a, b) .

El método de acertar o fallar es muy poco preciso y no se suele usar. Además el método de Monte Carlo crudo tiene aplicaciones más amplias al no estar restringido a intervalos ni a funciones positivas unidimensionales.

IV.6 Técnicas de reducción de la varianza

El objetivo de estas técnicas es incrementar la eficiencia estadística del análisis de simulación.

El medio para lograrlo es reduciendo la varianza de las variables de salida sin modificar la media de la estimación. De esta forma se obtiene mayor precisión para el mismo número de datos, o la precisión deseada con menos pasadas de simulación.

Antes de ver las técnicas de reducción de la varianza hay algunas observaciones que hacer sobre ellas:

- Los métodos de reducción de la varianza dependen del modelo en estudio, de modo que su comportamiento puede ser muy diferente en distintos modelos.
- Normalmente no es posible saber de antemano cuánto se va a poder reducir la varianza, o incluso, si se va a poder reducir.
- Algunas técnicas pueden aumentar el coste computacional, por lo que hay que buscar un equilibrio entre la mejora que pueden proporcionar y el coste que conlleva utilizarlas.

Las técnicas utilizadas más habitualmente para reducir la varianza de un estimador son las siguientes:

- Muestreo correlado.
- Utilización de variables de control.
- Utilización de variables antitéticas.
- Condicionamiento
- Muestreo estratificado.
- Muestreo por importancia

Vamos a ver una introducción a todas ellas, para ampliar el estudio puede consultarse cualquier libro de teoría de muestras o de simulación avanzado(ver [Law, 2000] o [Ríos-Insúa, 1997])

IV.6.1 Muestreo correlado (números aleatorios comunes)

El objetivo de esta técnica es comparar dos o más configuraciones alternativas para el sistema “bajo condiciones de experimentación similares”. Por lo tanto, se utiliza cuando se quiere comparar la variable de salida con dos configuraciones diferentes.

La idea básica es que se quiere comparar la variable de salida para dos configuraciones, y por lo tanto si X_{1j} y X_{2j} son las observaciones para la 1ª y 2ª configuración, respectivamente, en el tiempo j , lo que queremos estimar es

$$\xi = \mu_1 - \mu_2 = E(X_{1j}) - E(X_{2j})$$

Sea la variable $Z_j = X_{1j} - X_{2j}$, cuya esperanza es el valor que se desea estimar, $E(Z_j) = \xi$, y sea la media muestral de estas variable, $\overline{Z(n)} = \sum_{j=1}^n Z_j / n$, que es un estimador insesgado de ξ .

La varianza de este estimador es

$$V(\overline{Z(n)}) = \frac{V(Z_j)}{n} = \frac{V(X_{1j}) + V(X_{2j}) - 2Cov(X_{1j}, X_{2j})}{n}$$

Si las variables obtenidas con distintas configuraciones se obtienen de forma independiente, la covarianza es cero, pero se puede disminuir la varianza del estimador si se logra una covarianza positiva entre ambas. Para lograrlo, el muestreo correlado propone utilizar los mismos valores $U(0,1)$ para simular cada una de las configuraciones a lo largo del tiempo.

El método es eficaz y razonable, sin embargo, hay que tener cuidado al aplicarlo con la sincronización, pues puede no ser suficiente con poner la misma semilla para los números aleatorios. Una forma de buscar esa sincronización es utilizar una semilla diferente para cada variable aleatoria del modelo, con el fin de que aunque alguna de ellas pierda la sincronía el resto no lo haga.

IV.6.2 Variables antitéticas

Esta técnica se utiliza para reducir la varianza del estimador de la media, pero sin comparaciones entre distintas configuraciones.

El objetivo es inducir una correlación negativa entre las sucesivas simulaciones o replicaciones del modelo que se llevan a cabo para obtener una muestra con la que obtener el estimador.

Supongamos que X_1 y X_2 es una muestra de tamaño 2 de la variable respuesta. El estimador de la media de la variable será la media muestral, $\overline{X} = (X_1 + X_2)/2$ y su varianza será,

$$V(\overline{X}) = 1/4(V(X_1) + V(X_2) + 2Cov(X_1, X_2))$$

Obsérvese que si las variables son independientes, la covarianza es cero, pero que el valor de la varianza mejoraría si la covarianza fuera negativa. Por lo tanto, el objeto de esta técnica es intentar lograr una covarianza negativa entre dos replicaciones del mismo modelo. Para ello, el método propone que en una replicación se utilicen unos valores de la uniforme y en la otra los valores complementarios, es decir, 1 menos los anteriores.

Así si U_1, \dots, U_n son variables aleatorias con distribución uniforme en $(0,1)$, las variables $1-U_1, \dots, 1-U_n$ también son uniformes en $(0,1)$. Y es más, si X_1, \dots, X_n son variables obtenidas mediante la transformada inversa a partir de U_1, \dots, U_n , y X_1', \dots, X_n' son las obtenidas a partir de $1-U_1, \dots, 1-U_n$, entonces las variables X_j y X_j' están correladas negativamente.

De aquí se deduce que si en una simulación se utilizan los valores de una uniforme y en la siguiente sus complementarios las variables estarán correladas negativamente, mejorando el valor de la varianza del estimador. Sin embargo, no siempre es cierto ya que aunque las variables de entrada estén correladas negativamente, si el sistema es complejo las de salida pueden no estarlo. Además, el resultado está demostrado para variables de entrada generadas por el método de la transformada inversa, pero no siempre éste es el método que se utiliza para generar una variable aleatoria.

IV.6.3 Variables de control

El objetivo de esta técnica es utilizar la correlación entre ciertas variables aleatorias para conseguir reducir la varianza.

Supongamos que X es la variable de salida del modelo. Por ejemplo, en un sistema de colas podría ser el tiempo de espera en cola de los primeros 100 clientes, y se desea estimar $\mu = E(X)$.

Sea Y una variable aleatoria que aparece en el proceso de simulación, correlada con X (positiva o negativamente), y con esperanza $\nu = E(Y)$ conocida. Para el mismo ejemplo podría ser Y : tiempos de servicio de los primeros 99 clientes que terminan su servicio, cuya media es conocida puesto que el tiempo de servicio es una variable de entrada con media conocida.

Entonces el valor observado para la variable Y (valor que puede ser $Y > \nu$ o $Y < \nu$) permite ajustar el valor de X , y por lo tanto, se dice que Y es una *variable de control* para X .

Veamos cómo utilizar esta variable de control para mejorar la eficiencia del estimador. Para ello se define el siguiente *estimador de control*:

$$X_C = X - a(Y - \nu)$$

Obsérvese que este estimador es un estimador insesgado para μ , ya que

$$E(X_C) = E(X - a(Y - \nu)) = E(X) - a(E(Y) - \nu) = \mu - a(\nu - \nu) = \mu,$$

y su varianza es $V(X_C) = V(X) + a^2V(Y) - 2aCov(X, Y)$

Así pues, para que la varianza de este estimador sea mejor que la anterior ha de verificarse que $a^2V(Y) - 2aCov(X, Y) < 0$, y por lo tanto lo que hay que determinar es un valor de a para que esto sea cierto. En concreto, el mejor valor para la constante, el que da la menor varianza, es $a^* = -\frac{Cov(X, Y)}{V(y)}$.

Precisamente, el problema que puede tener este método es determinar la variable de control y el valor de a , que además requiere de una estimación previa de la covarianza entre las dos variables. Además, este método es específico para cada modelo, con lo cuál su automatización en el software de simulación comercial existente resulta prácticamente inviable.

IV.6.4 Condicionamiento

La técnica de condicionamiento para reducir la varianza del estimador intenta aprovechar alguna propiedad especial del modelo para remplazar un valor estimado por un valor analítico exacto. Al eliminar esta fuente de variabilidad es de esperar que la variable aleatoria de salida sea más estable, aunque no esté absolutamente garantizado. Esta técnica tiene su origen en el método de Monte Carlo condicional introducido por Trotter y Tukey (1956) y desarrollado por Hammersley and Handscomb (1964).

Supongamos que X es la variable de salida de nuestro modelo, cuya media μ desea ser estimada. Supongamos que existe otra variable aleatoria Z tal que para un valor dado de ésta, z , es posible calcular analíticamente el valor de la esperanza condicionada de X por este valor, es decir, $E[X / Z = z]$ es un valor conocido. Entonces un estimador insesgado para μ se puede obtener a partir de la esperanza en Z de la esperanza condicionada de X a los valores de Z , es decir, $\mu = E[X] = E_Z[E(X / Z)]$. Como ejemplo de lo que sería este estimador, si Z fuera discreta aunque de distribución desconocida el estimador por condicionamiento sería $E_Z[E(X / Z)] = \sum_z E(X / Z = z)p(z)$, que sería centrado. Respecto a la reducción de la varianza se verifica la siguiente relación ya que la varianza es una función no negativa

$$Var_Z[E(X / Z)] = Var[X] - E_Z[Var(X / Z)] \leq Var[X]$$

Así el procedimiento sería muestrear para obtener valores de la variable Z y para cada uno de ellos calcular analíticamente el valor de la esperanza condicionada de X por ese valor. Es decir, se realiza el experimento de simulación para observar la variable Z obteniéndose las observaciones $\{z_1, \dots, z_n\}$ y a partir de ahí se calcula el estimador $\sum_{i=1}^n \frac{1}{n} E[X / Z = z_i]$.

Una aplicación de este método puede verse en Carter and Ignall (1975). En este caso se querían analizar políticas de atención del servicio de extinción de incendios del Bronx, comparando para ello el tiempo medio de respuesta cuando se produce un incendio considerado “serio” (en el que pueden peligrar vidas humanas). Datos históricos revelaban que 1 de cada 30 incendios es serio. Y por lo tanto para hacer la simulación había que simular 30 veces más situaciones que las que se querían considerar para la variable de salida. Si embargo, el modelo tenía tal estructura que conociendo la situación de los equipos de extinción en un momento dado, era posible determinar analíticamente con exactitud el tiempo medio de respuesta si en ese momento se produjera un incendio serio. Así el procedimiento consistió en hacer la simulación, pero interrumpir periódicamente la simulación para observar el estado del sistema y calcular y almacenar el valor del tiempo medio de respuesta bajo esas condiciones si se produjera un incendio serio en ese momento (aunque realmente no estuviera sucediendo). El estimador final fue la media de estos tiempos de respuesta esperados condicionados e incluía muchos más términos que el número de incendios serios que realmente habían sido simulados. Con este procedimiento se redujo en un 95 % la varianza del estimador aunque con un mayor esfuerzo computacional. Para el mismo esfuerzo computacional la reducción fue de un 92 %.

Sin embargo en este método no siempre se puede asegurar que reduzca la varianza en tales dimensiones, o incluso que se llegue a reducirla, dada la correlación que puede existir entre los valores condicionados de X y que no ha sido tenida en cuenta.

IV.6.5 Muestreo estratificado

Esta técnica para reducir la varianza es una técnica no específica de la simulación sino que proviene de la idea de estratificación de la teoría de muestras (ver Thomson, 1992).

Supongamos que deseamos estimar $\mu = E[h(X)]$ para una variable aleatoria X con función de densidad $f(x)$, $x \in D$. Es posible hacer una partición del soporte D en k subconjuntos o *estratos* D_i y definir $\mu_i = \int_{D_i} h(x)f(x)dx$. Entonces, se tiene que

$$\mu = \int_D h(x)f(x)dx = \sum_{i=1}^k \int_{D_i} h(x)f(x)dx = \sum_{i=1}^k \mu_i$$

Así pues, el procedimiento sería dividir el soporte en estratos y estimar el valor en cada estrato y después sumarlos. Se comprueba que si la estratificación se ha hecho adecuadamente, se obtiene una reducción en la varianza al tener menor variabilidad en los estratos que en el soporte completo. El efecto de la reducción se alcanzará concentrando puntos muestrales en estratos que sean más importantes.

La mayor dificultad de este método está en cómo elegir los estratos para lograr reducir la varianza y el tamaño de la muestra en cada estrato. Por otra parte, una hipótesis implícita en la estratificación es que resulta posible generar muestras directamente de los espacios muestrales condicionados asociados a cada estrato, lo cuál no siempre es posible.

IV.6.6 Muestreo por importancia

Este método se basa en los conceptos de estratificación y ponderación de la teoría de muestras (ver Thomson, 1992). Supongamos que queremos estimar $\mu = E[h(X)]$, donde la variable aleatoria X tiene función de densidad $f(x)$, $x \in D \subset \mathbb{R}^m$. En tal caso, el problema se plantea como estimar

$$\mu = E[h(X)] = \int_D h(x)f(x)dx$$

La idea básica de este método consiste en observar que para cualquier otra variable aleatoria Z con función de densidad g con el mismo dominio D , denominada *distribución de importancia*, se puede expresar el estimador como

$$\mu = \int_D \frac{h(z)f(z)}{g(z)} g(z)dz = E_Z \left(\frac{h(z)f(z)}{g(z)} \right)$$

Así se puede estimar el parámetro muestreando los valores z_i de Z y utilizar como estimador

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \frac{h(z_i)f(z_i)}{g(z_i)}$$

que es un estimador insesgado. Básicamente, este estimador es una media de los valores $h(z_i)$ ponderados con pesos f/g . La varianza de este estimador puede llegar a ser mucho menor que la de X si se toma la distribución de importancia con forma similar a hf de modo que el cociente sea casi constante.

Esta técnica concentra la distribución en los puntos que se consideran de mayor importancia, de modo que si supiésemos de antemano que algunos valores son más importantes que otros en la determinación del parámetro, deseáramos seleccionar estos valores con mayor frecuencia de la que la distribución les da en sí. Sin embargo, esto puede ser difícil de conseguir y suele ser conveniente llevar a cabo una simulación piloto de tamaño pequeño para determinar si se produce reducción de la varianza.

Por otra parte, como ocurre con la mayoría de las técnicas de reducción de la varianza, el muestreo por importancia depende fuertemente del modelo, por lo que resulta interesante sugerir distribuciones de importancia adecuadas para clases de problemas relevantes. Aunque es difícil hacerlo, se ha avanzado en algunas áreas (ver Geweke (1989) para la aplicación a algunos modelos econométricos).

IV.7 Diseño de experimentos

Por último dentro del análisis de resultados vamos a ver una breve introducción de cómo diseñar los experimentos para poder comparar distintas alternativas sobre la configuración o las estrategias.

Cuando se quieren comparar diferentes alternativas, suele haber ciertos factores controlables (cuantitativos o cualitativos) que operan a distintos niveles (valores) que definen las alternativas. En un estudio de simulación interesa identificar el efecto de cada factor al variar en sus distintos niveles, la interacción entre los factores, etc. a la hora de tomar una decisión.

La cuestión es qué pruebas o replicaciones hay que hacer para identificar esos efectos. Para dar respuesta a esta pregunta la técnica estadística apropiada es el *diseño de experimentos*.

Si bien es cierto que esta técnica suele resultar más relevante cuando los experimentos no son simulados por la dificultad de llevarlos a cabo, también en simulación cuando hay muchos factores o muchos posibles valores o las simulaciones son muy largas, es deseable diseñar los experimentos evitando los diseños completos, es decir, evitando probar todas las combinaciones posibles de factores y niveles para poder comparar. Obsérvese que con k factores y 2 niveles cada uno y realizando m replicaciones de cada combinación, resultan $m2^k$ simulaciones.

Existen diversos diseños que pueden aplicarse como la variación de un factor cada vez, los diseños factoriales globales, fraccionales, superficie de respuesta, ...

Para un estudio de estos diseños se remite al lector a un libro de estadística (como por ejemplo [Peña, 98]) o uno específico de diseño de experimentos ([Montgomery, 91]). Sin embargo, a continuación se va a presentar una breve introducción a los diseños más habituales.

IV.7.1 Principios del diseño experimental

El objetivo de un experimento es estudiar el efecto que sobre una variable respuesta tienen un conjunto de otras variables denominadas *variables experimentales*, *factores* o *tratamientos*.

Supondremos que la variable respuesta es continua y que los factores se fijan durante el experimento a ciertos niveles determinados. El experimento consiste en seleccionar distintas

unidades experimentales, fijar los valores de los factores a distintos niveles y observar el valor de la variables respuesta en cada unidad. El número total de datos es el *tamaño del experimento*.

En cualquier experimento en que se investiga el efecto de un factor hay un gran número de otras variables que pueden influir en los resultados. Básicamente hay tres caminos para eliminar el efecto de una variable:

1. mantenerla fija durante toda la realización del experimento,
2. reorganizar la estructura del experimento para las comparaciones de interés se efectúen para valores fijos de esta variable
3. aleatorizar su aparición en los tratamientos

Los experimentos científicos suelen usar los dos primeros para eliminar el efecto de aquellas variables que son controladas por el experimentador, y el tercero para eliminar el efecto de variables fuera de nuestro control y de poca influencia esperada (se englobarán dentro del error experimental).

En general, cuando se analizan resultados de simulación no suele utilizarse la aleatorización ya que en el modelo no suelen existir variables fuera de control, sin embargo, se va a explicar el principio de aleatorización sobre todo para tenerlo en cuenta a la hora de tomar datos y observaciones para crear el modelo, y a la hora de validarlo comparando con la realidad.

El principio de aleatorización

Este principio requiere que todos los factores no controlados por el experimentador y que puedan influir en los resultados se asignen al azar a las observaciones. Por ejemplo, si al comparar las ventas de un producto en un establecimiento en dos días distintos de la semana sospechamos que puede influir el empleado que esté en esa sección que pueden ser dos distintos, deben tomarse observaciones para esos días con el vendedor asignado de forma aleatoria a las observaciones (aleatoria, es decir, con orden aleatorio también, para evitar efectos de aprendizaje, etc.).

La aleatorización es fundamental ya que previene la existencia de sesgos, evita la dependencia entre observaciones y confirma la validez de los procedimientos estadísticos más comunes.

De este principio se derivan los denominados *contrastes de permutaciones* para analizar la influencia de un factor en presencia de otra variable:

Supóngase que para el ejemplo anterior se deciden observar 5 días de un tipo (D1) y 5 de otro (D2). Se asignan de forma aleatoria los dos empleados a cada día de forma que siempre haya 5 días de cada uno en los 10, y se observa la diferencia de medias en la variable respuesta

de ambos tipos de día: $\bar{x}_{D_1} - \bar{x}_{D_2}$. Si el día observado no influyera (que es lo que se pretende determinar) daría igual que las observaciones fueran de un día u otro, así pues podríamos considerar que de las 10 observaciones cualesquiera 5 son de un día y el resto de otro. El número de asignaciones posibles de los tipos de día a las 10 observaciones es $\binom{10}{5} = 252$, y si se calculan las 252 diferencias que se obtienen con las distintas asignaciones y el valor concreto que se tiene está entre el $100\alpha/2$ % de las máximas o de las mínimas se considera que un valor tan anormalmente alto o bajo sólo puede ser debido a que no son iguales las medias.

Los test de permutaciones fueron introducidos por Fisher (1953), y Scheffé (1959) demostró que los contrastes t y F (análisis de la varianza para igualdad de medias) son una buena aproximación incluso sin la hipótesis de normalidad que se les exige.

La repetición del experimento

Como ya se ha comentado, para reducir la varianza de la estimación de una media que es σ^2/n , se puede reducir el valor de σ^2 o aumentar n que conlleva replicar el experimento. Aunque este tema ya ha sido tratado, hay que indicar la diferencia entre repetir cada observación y repetir el experimento. Si se toman dos medidas seguidas en cada condición experimental se tiene cada observación repetida dos veces, pero no necesariamente un experimento replicado. Conviene realizar el experimento una vez y luego repetirlo desde el principio para evitar subestimar la variabilidad.

La homogeneidad estadística de las comparaciones: diseños factoriales

Para comparar el efecto de un factor en la variable respuesta es claro que se debe comparar con distintos niveles pero actuando éstos en condiciones homogéneas. Existen dos enfoques para intentar lograr esa homogeneidad:

- a) diseños clásicos: eliminar del experimento las demás variables manteniéndolas fijas a niveles constantes y variar sólo la que se trata de comparar
- b) diseños factoriales: introducir en la experimentación todas las variables que pueden influir y estimar los efectos promediando situaciones homogéneas. Así para el ejemplo de la venta del producto se obtendrían datos para cada nivel del factor día con los dos empleados, siendo una tabla de doble entrada.

De los dos enfoques el primero tiene como inconveniente principal que no tiene en cuenta la interacción entre los factores. Por ejemplo, supongamos que se desea investigar si la

temperatura y la concentración influyen en el rendimiento de un proceso. Con el diseño clásico fijaríamos la temperatura a un valor de referencia y modificaríamos el valor de la concentración para ver su efecto; después haríamos lo mismo pero fijando la concentración a un valor y variaríamos la temperatura. De esta forma es posible llegar a conclusiones tales como que aumentar la temperatura disminuye el rendimiento y aumentar la concentración también, y sin embargo, es posible que aumentando temperatura y concentración a la vez se logre un mayor rendimiento. Esto ocurre porque los efectos pueden no ser aditivos, es decir, el modelo puede ser del tipo $Rend = b_1 Conc + b_2 Temp + b_3 Conc \cdot Temp$, y sin embargo, con el diseño clásico se supone que el coeficiente b_3 es cero.

El concepto de bloque

Una variable o factor cuyo efecto sobre la respuesta no es directamente de interés pero que se utiliza para obtener comparaciones homogéneas se denomina *variable bloque*. En el ejemplo, de la venta del producto sería el empleado si lo que realmente nos interesa saber es sólo el efecto del tipo de día. La principal diferencia en el tratamiento de un factor y de una variable bloque es que, en general, se supone que hay independencia entre los factores y las variables bloque.

IV.7.2 Diseño con un factor

Éste es el diseño básico en el cuál no se presenta más que un único factor que opera a I niveles, de modo que se dispone de una muestra de tamaño n tal que n_i han sido obtenidas con el factor fijado al nivel i , siendo y_{ij} estas observaciones.

El modelo que supone este factor es $y_{ij} = \mu_i + u_{ij}$, siendo u_{ij} de distribución $N(0, \sigma)$ independientes (se puede comprobar representando los valores para cada nivel y ver que siguen una distribución normal).

El procedimiento es en primer lugar estimar los parámetros del modelo: las μ_i y σ^2 ; en segundo lugar realizar el contraste de igualdad de medias para ver si el factor influye en la variable respuesta, es decir, contrastar si es cierto que $\mu_1 = \dots = \mu_I = \mu$. Después comprobar con los residuos las hipótesis del modelo, y por último, si se ha rechazado que el factor no influye hacer comparaciones de las medias.

Los estimadores de los parámetros del factor son $\hat{\mu}_i = \bar{y}_i = \frac{\sum_j y_{ij}}{n_i}$, y definiendo los residuos como $e_{ij} = y_{ij} - \bar{y}_i$, el estimador de la varianza es $\hat{\sigma}^2 = \frac{\sum_{ij} e_{ij}^2}{n}$.

A continuación se plantea el contraste de igualdad de medias:

$H_0 : \mu_1 = \dots = \mu_I = \mu$ frente a la hipótesis de que hay alguna distinta. Para hacer este contraste se compara la variabilidad explicada por los grupos (VE) con la no explicada o residual (VNE), de modo que si es muy grande la explicada respecto a la que no entonces se rechaza la hipótesis nula y se acepta que el factor influye en la variable respuesta. Este análisis se conoce como *análisis de la varianza* y para llevarlo a cabo se utiliza una tabla denominada ANOVA (ADEVA en castellano):

Fuentes variación	Suma de cuadrados	Grados de libertad	Varianzas
Entre grupos (VE)	$\sum_i n_i (\bar{y}_i - \bar{y}_{..})^2$	$I - 1$	\hat{s}_e^2
Interna, no explicada o residual (VNE)	$\sum_{i,j} (y_{ij} - \bar{y}_i)^2$	$n - I$	\hat{s}_R^2
Total	$\sum_{i,j} (y_{ij} - \bar{y}_{..})^2$	$n - 1$	\hat{s}_y^2

donde la última columna es el cociente de las dos anteriores.

Si la hipótesis nula fuera cierta el estadístico $F_{(I-1),(n-I)} = \frac{\hat{s}_e^2}{\hat{s}_R^2}$ seguiría una distribución F de Fisher-Snedecor con esos grados de libertad indicados, de modo que se rechazará la hipótesis si este estadístico toma un valor suficientemente alto.

Así si se acepta la hipótesis nula, se acepta la igualdad de medias y por lo tanto que el factor no influye. Si se rechaza tiene interés estimar las diferencias entre los grupos. Para ello se puede hacer un intervalo de confianza para la diferencia de medias de los grupos dos a dos o hacer un contraste de igualdad de medias dos a dos. Sin embargo, hay que tener cuidado ya que la aplicación reiterada de ese contraste (para I niveles serían $\binom{I}{2}$ contrastes) puede llevar a que

aparezcan diferencias entre grupos cuando realmente no las hay. Para evitarlo el método de Bonferroni propone que cada contraste debería ser hecho al nivel de significación α/c , donde α es el nivel de significación total que se desea y c el número de comparaciones que se van a realizar.

IV.7.3 Diseño en bloques aleatorizados

Este diseño se utiliza para analizar el efecto de un factor en presencia de una variable bloque. Se tienen I niveles del factor, J valores de la variable bloque y se tienen $I \cdot J$ observaciones de la variable respuesta obtenidas cada una fijando un nivel del factor y un valor del bloque, denotadas por y_{ij} .

El modelo en bloques aleatorizados es $y_{ij} = \mu + \alpha_i + \beta_j + u_{ij}$, $i = 1, \dots, I$, $j = 1, \dots, J$, donde u_{ij} es una variable $N(0, \sigma)$ y sus valores independientes.

El modelo descompone la respuesta como suma de un efecto global μ , un efecto incremental del factor principal α_i (al ser incremental, se supone $\sum_i \alpha_i = 0$), un efecto incremental del factor secundario o bloque β_j (de nuevo $\sum_j \beta_j = 0$) y un efecto aleatorio u_{ij} que sería el denominado error experimental debido a las restantes causas de variabilidad.

El objetivo primero del análisis es determinar si los efectos del factor principal son nulos, y si se decide que no entonces puede ser de interés analizar las diferencias.

El primer paso es estimar los valores de los parámetros que serán un total de $I + J$, donde se incluyen el valor de μ , los α_i , los β_j y σ^2 .

El estimador de μ es la media de todas las observaciones, que denominaremos $\bar{y}_{..}$; los estimadores de los efectos del factor principal se obtienen como la media de las observaciones para ese nivel del factor menos la media de todas las observaciones, es decir,

$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} = \frac{1}{J} \sum_j y_{ij} - \bar{y}_{..}$; análogamente para estimar los efectos del bloque se tendría

$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$; por último, para la varianza del error, se obtienen los residuos

$e_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$ y se estima como $\hat{\sigma}^2 = \frac{1}{n} \sum_{i,j} e_{ij}^2$.

Una vez estimados los coeficientes se procede a contrastar si se pueden considerar significativamente distintos de cero y poder concluir que realmente el factor influye. Además

interesa saber si el bloque influye, ya que si no fuera así, se podría optar por un modelo más sencillo, como el de la sección anterior, que no distinguiera por el bloque. Para hacer estos contrastes se realiza el *análisis de la varianza*. La lógica de este análisis es en primer lugar descomponer la variabilidad total de los datos en sus fuentes de variación, y comparar el término debido a cada fuente con un patrón común, la varianza residual, para juzgar respecto a su importancia.

La tabla ANOVA en este caso será de la siguiente forma

Fuente	Suma cuadrados	G. libertad	Varianzas
Entre tratamientos (distintas α)	$J \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = J \sum_i \hat{\alpha}_i^2$	$I - 1$	\hat{s}_α^2
Entre bloques (distintas β)	$I \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = I \sum_j \hat{\beta}_j^2$	$J - 1$	\hat{s}_β^2
Residual	$\sum_{i,j} e_{ij}^2 = \sum_{i,j} (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$	$(I - 1)(J - 1)$	\hat{s}_R^2
Total	$\sum_{i,j} (y_{ij} - \bar{y}_{..})^2$	$n - 1$	\hat{s}_y^2

A partir de esta tabla se realizan los contrastes. El contraste principal es que los tratamientos son idénticos:

$$H_0 : \alpha_i = 0; \quad \forall i$$

Se efectúa con el estadístico $F_{(I-1),(I-1)(J-1)} = \frac{\hat{s}_\alpha^2}{\hat{s}_R^2}$, que si la hipótesis nula es cierta sería una F de Fisher-Snedecor con los grados de libertad indicados, de modo que si el valor del estadístico es alto será rechazada la hipótesis de que el factor no influye.

Para los bloques el contraste es $H_0 : \beta_j = 0; \quad \forall j$ y el estadístico es

$$F_{(J-1),(I-1)(J-1)} = \frac{\hat{s}_\beta^2}{\hat{s}_R^2}, \text{ y es un contraste independiente del anterior.}$$

La eficacia del diseño depende de los efectos de los bloques, pero en cualquier caso nunca es peor que si se tratara como un diseño completamente aleatorio como el de la sección anterior.

Si del análisis de la varianza se deriva que el factor influye entonces se pasa a realizar comparaciones entre las medias de los grupos determinados por los niveles del factor, tal y cómo se comentó en la sección anterior.

Respecto a la validación de modelo, el principal error que se puede estar cometiendo es que se está despreciando la interacción entre el factor y el bloque, y podría ser que existiera, es decir, el modelo podría ser de la forma:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ij}$$

Este modelo tiene demasiados parámetros, por lo que Tukey propuso tomar $(\alpha\beta)_{ij} = \gamma\alpha_i\beta_j$ que sólo añade un parámetro.

La interacción se detecta fundamentalmente porque si ésta existe y no ha sido considerada en el modelo, los residuos no tendrán esperanza nula, y gráficamente si se representa en el eje de abscisas el valor previsto y en el de ordenadas el residuo se podrá ver una curvatura.

IV.7.4 Modelos factoriales: con dos factores y con interacción

Este modelo se aplica cuando las dos variables a estudiar tienen la misma importancia a priori y se admite la interacción desde un principio.

El modelo sin replicación

El modelo que se considera es el siguiente:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ij}$$

y el número de parámetros a estimar son $IJ + 1$ teniendo en cuenta las restricciones $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$. Como se puede observar hay más parámetros que observaciones, luego no pueden ser todos estimados. La opción es no estimar la varianza, y estimar los parámetros como en el caso anterior y para las interacciones estimarlas como los residuos, es decir, $(\alpha\beta)_{ij} = y_{ij} - \bar{y}_i - \bar{y}_j - \bar{y}_{..}$. Sin embargo al no disponer de residuos no es posible contrastar el modelo, salvo que se consideren las interacciones nulas, o que se replique el experimento.

El modelo con replicación

Suponiendo que se hacen K replicaciones para cada combinación de niveles de los dos factores se dispone de IJK observaciones y el modelo considerado ahora sería:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ijk}$$

Aunque el número de parámetros a ser estimados sigue siendo $IJ + 1$ ahora hay más observaciones. Los estimadores en este caso serán

$$\begin{aligned}\hat{\mu} &= \bar{y}_{...} \\ \hat{\alpha}_i &= \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} \\ (\alpha\beta)_{ij} &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\end{aligned}$$

Los residuos en este caso serán: $e_{ijk} = y_{ijk} - \bar{y}_{ij.}$. La tabla ANOVA será en tal caso la siguiente

Fuente	Suma cuadrados	G. libertad	Varianzas
Factor α	$JK \sum_i \hat{\alpha}_i^2$	$I - 1$	\hat{s}_α^2
Factor β	$IK \sum_j \hat{\beta}_j^2$	$J - 1$	\hat{s}_β^2
Interacción	$K \sum_{i,j} (\alpha\beta)_{ij}^2$	$(I - 1)(J - 1)$	$\hat{s}_{\alpha\beta}^2$
Residual	$\sum_{i,j,k} e_{ijk}^2$	$IJ(K - 1)$	\hat{s}_R^2
Total	$\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2$	$IJK - 1$	\hat{s}_y^2

Los tests resultantes y estadísticos serán los siguientes:

- 1) $H_0 : \alpha_i = 0, \forall i;$ estadístico $F_{(I-1),IJ(K-1)} = \frac{\hat{s}_\alpha^2}{\hat{s}_R^2}$
- 2) $H_0 : \beta_j = 0, \forall j;$ estadístico $F_{(J-1),IJ(K-1)} = \frac{\hat{s}_\beta^2}{\hat{s}_R^2}$
- 3) $H_0 : (\alpha\beta)_{ij} = 0, \forall i, j;$ estadístico $F_{(I-1)(J-1),IJ(K-1)} = \frac{\hat{s}_{\alpha\beta}^2}{\hat{s}_R^2}$

IV.7.5 Modelos factoriales con más de dos factores

El modelo con tres factores sería de la forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + u_{ijk}$$

En este modelo además de las interacciones de segundo orden para cada par de variables, existe una interacción de tercer orden. En general, los efectos principales son mayores que los de segundo orden, y éstos mayores que los de tercer orden.

De nuevo, como ocurría en el caso de dos factores, hay más parámetros que observaciones, luego no será posible estimar todos ellos, siendo la varianza del ruido el que se deja sin estimar. Sin embargo, por esa misma razón los contrastes de la tabla ANOVA sólo pueden realizarse si se supone que la interacción de tercer orden es nula, o si se dispone de replicaciones.

El método de estimación y los contrastes son una extensión de los del caso de dos factores.

Estas ideas se extienden inmediatamente para modelos con cualquier número de factores, aunque para más de tres las interacciones superiores a tercer orden suelen suponerse nulas. La limitación principal de estos diseños es el número de observaciones que requieren, de modo que para más de cuatro factores no suelen usarse estos diseños completos y se acude al concepto de *fracción*: suponiendo que las interacciones son nulas a partir de un orden se selecciona una parte del diseño completo que permita estimar los efectos principales y las interacciones de orden bajo. La siguiente sección muestra uno de estos diseños fraccionales.

IV.7.6 Cuadrados latinos

Este modelo se puede aplicar en las siguientes condiciones:

- a) existen tres factores (alguno puede ser una variable bloque);
- b) el número de niveles para cada factor es el mismo;
- c) no se espera interacción entre los factores.

La ventaja de este modelo es que el número de observaciones se reduce mucho, de modo que si N es el número de niveles de cada factor, el diseño completo requiere N^3 observaciones mientras que el cuadrado latino sólo N^2 , es decir, se trata de un diseño fraccional.

La idea básica del cuadrado latino es aprovechar la simetría del diseño factorial seleccionando un conjunto de combinaciones de los niveles de los factores con la condición de que cada nivel de un factor aparezca una vez con cada uno de los niveles de los otros factores.

Para llevar a cabo la experimentación con un diseño de cuadrado latino en primer lugar se selecciona un cuadrado latino que corresponda al número de niveles con los que se está trabajando, y luego se aleatoriza el orden de las filas y columnas. Los posibles cuadrados latinos son los siguientes:

$$N = 3$$

<i>A</i>	<i>B</i>	<i>C</i>		<i>A</i>	<i>B</i>	<i>C</i>
<i>B</i>	<i>C</i>	<i>A</i>		<i>C</i>	<i>A</i>	<i>B</i>
<i>C</i>	<i>A</i>	<i>B</i>		<i>B</i>	<i>C</i>	<i>A</i>

$$N = 4$$

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>B</i>	<i>A</i>	<i>D</i>	<i>C</i>		<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>		<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>		<i>B</i>	<i>A</i>	<i>D</i>	<i>C</i>		<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>
<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>		<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>		<i>B</i>	<i>A</i>	<i>D</i>	<i>C</i>

$$N = 5$$

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>		<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>		<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>		<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>		<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>		<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>		<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>		<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>
<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>		<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>		<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>

$$N = 6$$

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<i>B</i>	<i>A</i>	<i>F</i>	<i>E</i>	<i>C</i>	<i>D</i>
<i>C</i>	<i>F</i>	<i>B</i>	<i>A</i>	<i>D</i>	<i>E</i>
<i>D</i>	<i>C</i>	<i>E</i>	<i>B</i>	<i>F</i>	<i>A</i>
<i>E</i>	<i>D</i>	<i>A</i>	<i>F</i>	<i>B</i>	<i>C</i>
<i>F</i>	<i>E</i>	<i>D</i>	<i>C</i>	<i>A</i>	<i>B</i>

Para ver los cuadrados latinos de mayor dimensión consultar Peña (1998).

Se asigna un factor a las filas, otro a las columnas y otro a las letras. El diseño es simétrico para ambas así que se pueden asignar como se desee, y en general, se hablará del efecto fila, efecto columna y efecto letra. A continuación se aleatoriza el orden de las columnas y el de las filas, resultando así las combinaciones de niveles de los tres factores que han de ser experimentadas.

Por ejemplo, si para comparar el efecto que tiene sobre el consumo en automóviles cuatro tratamientos de la gasolina, se seleccionan 4 automóviles y 4 conductores, se puede aplicar un diseño de cuadrado latino con $N = 4$ para eliminar en las comparaciones del efecto vehículo y conductor. Se asigna, por ejemplo, que las columnas sean los conductores (cada columna es un conductor), las filas el vehículo (cada fila un vehículo) y las letras el tratamiento.

Entonces se selecciona un cuadrado latino de orden 4, por ejemplo,

$$\begin{array}{cccc}
 A & B & C & D \\
 B & A & D & C \\
 C & D & A & B \\
 D & C & B & A
 \end{array}
 =
 \begin{array}{cccc}
 T1 & T2 & T3 & T4 \\
 T2 & T1 & T4 & T3 \\
 T3 & T4 & T1 & T2 \\
 T4 & T3 & T2 & T1
 \end{array}$$

Ahora, se aleatoriza a quién corresponde cada fila y cada columna. Así, si el orden resultante para las filas es (2,1,3,4) y para las columnas (3,4,2,1) el diseño resultante sería:

	<i>C3</i>	<i>C4</i>	<i>C2</i>	<i>C1</i>
<i>V2</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>
<i>V1</i>	<i>T2</i>	<i>T1</i>	<i>T4</i>	<i>T3</i>
<i>V3</i>	<i>T3</i>	<i>T4</i>	<i>T1</i>	<i>T2</i>
<i>V4</i>	<i>T4</i>	<i>T3</i>	<i>T2</i>	<i>T1</i>

Reordenando las filas y columnas, que es como se suele presentar el diseño,

	<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>
<i>V1</i>	<i>T3</i>	<i>T4</i>	<i>T2</i>	<i>T1</i>
<i>V2</i>	<i>T4</i>	<i>T3</i>	<i>T1</i>	<i>T2</i>
<i>V3</i>	<i>T2</i>	<i>T1</i>	<i>T3</i>	<i>T4</i>
<i>V4</i>	<i>T1</i>	<i>T2</i>	<i>T4</i>	<i>T3</i>

Cada celda representa un experimento que hacer con el conductor de la columna correspondiente, el vehículo de la fila correspondiente y el tratamiento de gasolina que se

indique en la celda. Obsérvese que cada nivel de un factor aparece combinado con todos los niveles de los demás factores. Por otra parte, un diseño completo exigiría $4^3 = 64$ experimentos mientras que este diseño requiere $4^2 = 16$ pruebas.

El modelo que se supone es el siguiente

$$y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_k + u_{ij(k)}, \quad i, j, k = 1, \dots, N$$

donde α_i es el efecto fila, β_j el efecto columna, γ_k el efecto letra en las distintas casillas y las perturbaciones $u_{ij(k)}$ son normales $(0, \sigma^2)$ independientes. La notación $ij(k)$ indica que el subíndice k depende de la casilla ij según el diseño, es decir, dados i y j , mirando en el cuadro, se tiene el valor de k . Se supone además que $\sum_1^N \alpha_i = \sum_1^N \beta_j = \sum_1^N \gamma_k = 0$.

La estimación de los parámetros sería la siguiente:

$$\begin{aligned} \hat{\mu} = \bar{y}_{...} &= \frac{\sum_{i,j} y_{ij}}{N^2} \\ \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}; \quad \bar{y}_{i..} &= \frac{\sum_j y_{ij}}{N} \\ \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}; \quad \bar{y}_{.j.} &= \frac{\sum_i y_{ij}}{N} \\ \hat{\gamma}_k = \bar{y}_{..k} - \bar{y}_{...}; \quad \bar{y}_{..k} &= \frac{\sum_{i,j} y_{ij(k)}}{N} \end{aligned}$$

Los residuos se obtienen como $e_{ij(k)} = y_{ij(k)} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k$, y la estimación centrada

de la varianza residual será $\hat{\sigma}^2 = \hat{S}_R^2 = \frac{\sum_{i,j} e_{ij(k)}^2}{(N-1)(N-2)}$.

Por último, el cuadro del análisis de la varianza sería

Fuente	Suma cuadrados	G. libertad	Varianzas
Efecto fila	$N \sum_i \hat{\alpha}_i^2$	$N - 1$	\hat{s}_α^2
Efecto columna	$N \sum_j \hat{\beta}_j^2$	$N - 1$	\hat{s}_β^2

Efecto letra	$N \sum_k \hat{\gamma}_k^2$	$N - 1$	\hat{s}_γ^2
Residual	$\sum_{i,j} e_{ij(k)}^2$	$(N - 1)(N - 2)$	\hat{s}_R^2
Total	$\sum_{i,j} (y_{ijk} - \bar{y}_{...})^2$	$N^2 - 1$	\hat{s}_y^2

Los tests resultantes y estadísticos serán los siguientes:

- 1) efecto fila $H_0 : \alpha_i = 0, \forall i;$ estadístico $F_{(N-1),(N-1)(N-2)} = \frac{\hat{s}_\alpha^2}{\hat{s}_R^2}$
- 2) efecto columna $H_0 : \beta_j = 0, \forall j;$ estadístico $F_{(N-1),(N-1)(N-2)} = \frac{\hat{s}_\beta^2}{\hat{s}_R^2}$
- 3) efecto letra $H_0 : \gamma_k = 0, \forall k;$ estadístico $F_{(N-1),(N-1)(N-2)} = \frac{\hat{s}_\gamma^2}{\hat{s}_R^2}$

Para cuatro variables existe una extensión de este diseño que son los cuadrados greco-latinos, que no se va a ver en esta introducción pero que puede ser consultado en cualquier libro de diseño de experimentos o estadística como Peña (1998).

IV.7.7 Modelos con efectos aleatorios: Diseños jerárquicos

El objetivo de los diseños anteriores es determinar el efecto que producen en la variable respuesta ciertos niveles prefijados de un factor, y es lo que se llama diseño de *efectos fijos*. Sin embargo, existen otros diseños donde estos niveles no se eligen, sino que al muestrear están presentes en la población, denominándose entonces diseños de *efectos aleatorios*. Los diseños de efectos fijos se utilizan cuando se desea medir cómo varía el nivel de la variable respuesta con los niveles de los factores, mientras que los de efectos aleatorios se utilizan para ver el efecto sobre la variabilidad de la respuesta. Por ejemplo, para saber si un el uso de determinados tipos de carburante es efectivo se realiza un diseño de efectos fijos, mientras que si se desea saber la variabilidad que el uso de distintos carburantes utilizados habitualmente introduce en el proceso, se utilizará un modelo de efectos aleatorios.

En los diseños con efectos aleatorios no se fijan los niveles del factor, de modo que el factor se considera una variable aleatoria cuyo efecto en la respuesta se puede apreciar en la variabilidad, y el objetivo del diseño es precisamente estimar la varianza del factor, es decir, no se estima α sino σ_α^2 , y el contraste principal no es $\alpha_i = 0$ sino $\sigma_\alpha^2 = 0$.

La formulación matemática del modelo es idéntica, pero en el diseño de efectos fijos los efectos representan la respuesta media y son parámetros fijos a estimar, mientras que en el de

efectos aleatorios son variables aleatorias normales de media 0 y varianza σ_α^2 , siendo esta varianza el parámetro a estimar.

La descomposición de la variabilidad y la tabla ANOVA se realiza igual en ambos modelos, y si no existe interacción los contrastes de que un factor no influye son idénticos. Sin embargo cuando existe interacción entre dos efectos, los contrastes cambian de modo que si en el de efectos fijos para ver el efecto de un factor se contrasta el valor de $\hat{s}_\alpha^2 / \hat{s}_R^2$, en el de efectos aleatorios se contrasta $\hat{s}_\alpha^2 / \hat{s}_{\alpha\beta}^2$.

Por otra parte, en el modelo de efectos aleatorios la estimación de los valores α_i, β_j, \dots no tiene sentido, por lo que no se plantea el problema de las comparaciones múltiples, ya que el objetivo es contrastar que los efectos existen, y en caso afirmativo, estimar las varianzas.

Un caso especial de este tipo de diseños es cuando los factores que intervienen están *anidados o jerarquizados*. Por ejemplo, si se desea estudiar la influencia en la opinión de los trabajadores del sector industrial en que trabajan y de la empresa en que están, no es posible cruzar todos los niveles de los factores. En primer lugar habrá que seleccionar sectores, después, para cada sector seleccionar empresas de éste, y dentro de cada empresa seleccionar trabajadores. Así el factor empresa puede medirse sólo a un nivel fijo del factor sector. Este tipo de diseños se denominan de *clasificación jerárquica o diseños anidados*, y en ellos al no existir cruces entre factores no puede existir interacción entre ellos.

El caso de diseños jerárquicos más importante es el modelo con dos factores aleatorios o *modelo de componentes de la varianza*. Supóngase que se quiere estimar la variabilidad en el rendimiento siendo un factor las máquinas y otro los operarios. Para ello se selecciona un conjunto de máquinas al azar (efecto aleatorio) y se toman varias muestras de la producción con distintos trabajadores en cada máquina. La diferencia fundamental con otros diseños es que no están todas las máquinas y que los trabajadores son distintos en cada máquina, no se cruzan.

El modelo que se establece es $y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + u_{ijk}$, donde y_{ijk} es la respuesta, α_i es el efecto aleatorio de la máquina que es un valor de una $N(0, \sigma_\alpha^2)$, $\beta_{j(i)}$ es el efecto del trabajador j con la máquina i cuya distribución es $N(0, \sigma_\beta)$ y u_{ijk} es el error experimental al actuar un trabajador en una máquina concreta y se considera $N(0, \sigma)$. Tomando varianzas se tiene $\sigma_y^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$.

Cuando se toma la misma cantidad de trabajadores por cada máquina los datos pueden disponerse en una tabla de doble entrada como en un diseño factorial con dos factores (máquina y operario) con replicación. Sin embargo, la interpretación es muy distinta ya que los operarios

en el diseño clásico eran los mismos para todas las máquinas, mientras que aquí son diferentes para cada máquina (el hecho de que sea por ejemplo la primera columna supone que es el primero en orden para cada máquina no el mismo operario).

La estimación en este caso, dado que están anidados los factores se ha de hacer de forma secuencial. Supongamos que se tienen I valores del factor principal, J del factor anidado (la misma cantidad en todos los valores del principal) y K repeticiones de cada uno.

Se estima en primer lugar la varianza residual como:

$$\hat{\sigma}^2 = \hat{s}_R^2 = \frac{\sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2}{IJ(K-1)}.$$

A continuación se estima la varianza debida al factor anidado

$$\hat{\sigma}_\beta^2 = \frac{\hat{s}_2^2 - \hat{s}_R^2}{K} \text{ donde } \hat{s}_2^2 = \frac{K \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..})^2}{I(J-1)}.$$

Por último, la variabilidad debida al factor principal se estima como

$$\hat{\sigma}_\alpha^2 = \frac{\hat{s}_1^2 - K\hat{\sigma}_\beta^2}{KJ} \text{ donde } \hat{s}_1^2 = \frac{JK \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}{I-1}.$$

La diagnosis del modelo se efectúa con los contrastes habituales.

IV.7.8 Diseños factoriales a dos niveles (2^k)

En la experimentación industrial es frecuente que se desee conocer el efecto sobre una variable respuesta de un gran número de factores, aunque en ocasiones no todos sean realmente relevantes. Los diseños factoriales a dos niveles que se van a ver, permiten experimentar con un gran número de factores y además son fáciles de fraccionar en bloques homogéneos favoreciendo la experimentación secuencial.

En primer lugar se va a ver como son estos diseños factoriales completos y después cómo fraccionarlos y aprovecharlo para hacer una experimentación secuencial.

Un diseño factorial a dos niveles de orden k , representado por 2^k supone que hay k factores con dos niveles cada uno. Veremos primero el caso más sencillo: 2^2 .

El diseño 2^2

Se supone que hay dos factores, A y B, con dos niveles. Se usará la notación de signos + y – para representar los dos niveles de cada factor (si son cuantitativos se usa el valor + para el mayor). Las observaciones correspondientes se disponen en general en una tabla, notándose por (o) a las observaciones con ambos factores en el nivel –, por (a) cuando el factor A está en su nivel + y el B en –, por (b) si es al revés, y por (ab) cuando ambos están en el nivel +. Así la tabla sería de la forma

A	B	Y
–	–	y_{11} (o)
+	–	y_{21} (a)
–	+	y_{12} (b)
+	+	y_{22} (ab)

El modelo factorial sería de la forma

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ij}; \quad i, j = 1, 2$$

Este modelo con las restricciones habituales de que sena efectos incrementales sobre el valor medio (las sumas de los parámetros nulas) y añadiendo las variables

$$X_k = \begin{cases} +1 & \text{si factor } k \text{ en nivel } + \\ -1 & \text{si factor } k \text{ en nivel } - \end{cases}, \text{ quedaría}$$

$$y_{ij} = \mu + \alpha_2 X_1 + \beta_2 X_2 + (\alpha\beta)_{22} X_1 X_2 + u_{ij}; \quad i, j = 1, 2$$

En este diseño hay un parámetro por cada factor y uno por interacción. Tal y como está expresado los coeficientes representan el incremento esperado en la respuesta cuando los factores se fijan al nivel +. Sin embargo, por conveniencia, se hablará de efecto de un factor como el efecto esperado en la respuesta cuando pasa de – a +. Estos efectos los vamos a denotar por la misma letra pero sin subíndice ya que no es necesario, verificándose las siguientes relaciones:

$$\alpha_2 = -\alpha_1 = \frac{1}{2}\alpha \quad \beta_2 = -\beta_1 = \frac{1}{2}\beta \quad (\alpha\beta)_{22} = \frac{1}{2}(\alpha\beta)$$

La estimación de estos nuevos parámetros sería de la forma

$$\hat{\mu} = \frac{1}{4}(o + a + b + ab) \quad \hat{\alpha} = \frac{1}{2}(a + ab - o - b) \quad \hat{\beta} = \frac{1}{2}(b + ab - o - a)$$

$$(\alpha\beta) = \frac{1}{2}(o + ab - a - b)$$

Una forma de recordar estas estimaciones es conocida como el *algoritmo de los signos*. Este algoritmo se aplica creando una tabla denominada *estándar* a partir de la tabla de observaciones. En ella se añade una columna AB que es el producto de las columnas A y B, y la estimación de cada parámetro se hace multiplicando la correspondiente columna de signos por las observaciones y dividiendo luego entre 2. Este algoritmo es aplicable también cuando hay más de dos factores. La tabla estándar entonces sería:

A	B	AB	Y
-	-	+	y_{11} (<i>o</i>)
+	-	-	y_{21} (<i>a</i>)
-	+	-	y_{12} (<i>b</i>)
+	+	+	y_{22} (<i>ab</i>)

En este análisis se estiman cuatro parámetros con cuatro observaciones, de modo que no es posible estimar el error experimental. Por lo tanto, para hacer intervalos de confianza o contrastes de hipótesis para los parámetros será necesario una estimación externa de σ^2 o replicar el experimento. Si se replica el experimento, se aplica el procedimiento visto de estimación a las medias y los residuos se calculan como $e_{ijk} = y_{ijk} - \bar{y}_{ij}$. A partir de ellos se estima la varianza residual con la suma de los cuadrados de los residuos dividida por $4(R - 1)$ donde R es el número de repeticiones.

Hay que tener una precaución con este modelo, y es que cuando las interacciones son altas, los efectos principales no describen de forma adecuada los resultados. Como regla general, cuando la interacción sea un tercio del promedio de los efectos principales la descripción mediante efectos principales e interacciones es inadecuada.

Una de las ventajas de este diseño frente a uno factorial clásico es que obtiene estimaciones más precisas, de menor varianza, o si se quiere, que para lograr la misma precisión es necesario un número menor de experimentos. En general, para k factores el diseño clásico requiere $(k + 1)/2$ veces el número de observaciones de un diseño factorial.

El modelo 2³

Ahora se tienen 3 factores con 2 niveles cada uno. El modelo es semejante al anterior pero ahora existirán 3 efectos principales, 3 interacciones de segundo orden y 1 interacción de tercer orden. El procedimiento es semejante al anterior siendo la tabla estándar de la forma

A	B	C	AB	AC	BC	ABC	Y
-	-	-	+	+	+	-	(o)
+	-	-	-	-	+	+	(a)
-	+	-	-	+	-	+	(b)
+	+	-	+	-	-	-	(ab)
-	-	+	+	-	-	+	(c)
+	-	+	-	+	-	-	(ac)
-	+	+	-	-	+	-	(bc)
+	+	+	+	+	+	+	(abc)

Para hacer la estimación del efecto de un factor o interacción se multiplica la columna de signos correspondiente por la columna de las observaciones, dividiendo después por $n/2 = 2^{k-1} = 2^2$. Sólo el estimador de la media va dividido por $n = 2^k$.

Una vez hechas las estimaciones para interpretarlas con intervalos de confianza o hacer contrastes es necesaria una estimación de la varianza residual. Hay básicamente tres formas de estimarla:

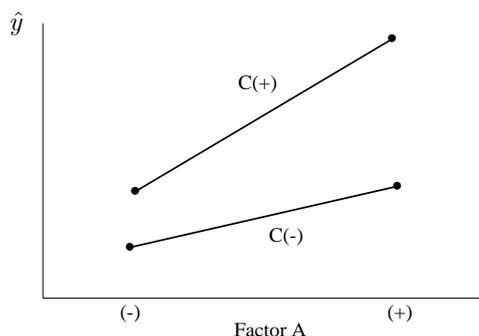
- a) mediante una estimación externa
- b) replicando el diseño: los efectos se estiman con las medias de los valores correspondientes y la varianza residual será la suma de los cuadrados de los residuos dividida entre $S(R - 1)$, siendo R el número de replicaciones; el contraste se basará

en el estadístico $\frac{\hat{\vartheta}}{\hat{s}_R / \sqrt{2R}}$ que sería una T de Student con $S(R - 1)$ grados de libertad

- c) considerando que las interacciones son nulas (se utiliza mucho en diseños con más factores considerando nulas las interacciones de tercer orden y superiores): si es así sus estimaciones deben ser valores aleatorios de una normal de media cero y varianza $\sigma^2/2$, y si no lo es y alguna interacción es no nula, la media cambiará pero no la varianza. Sea la mediana de las estimaciones de las interacciones $M(\hat{\vartheta}) = \text{mediana}(\hat{\vartheta}_5, \hat{\vartheta}_6, \hat{\vartheta}_7, \hat{\vartheta}_8)$. Un estimador robusto para la variabilidad es la MEDA o mediana de las desviaciones absolutas a la mediana: $MEDA = \text{mediana}_{i=5, \dots, 8} |\hat{\vartheta}_i - M(\hat{\vartheta})|$. Un estimador robusto para la desviación típica en poblaciones normales será $\hat{s}_{\vartheta} = \frac{MEDA}{0.675}$. Para realizar los contrastes de hipótesis se

considerará que el efecto principal de un factor existe si $|\hat{\vartheta}_j| \geq 2\hat{s}_{\vartheta}$, y una estimación del error experimental será $\hat{\sigma} = \hat{s}_{\vartheta}\sqrt{2}$.

En cuanto a decidir si una interacción puede considerarse nula es habitual decidirlo de forma gráfica o geométrica. Así si se desea ver si hay interacción entre el factor A y el C se hace la siguiente representación:



El valor \hat{y} representado es el valor esperado con los factores representados en la gráfica al nivel que se indica en cada punto. Por ejemplo para (-) en A y en C el valor esperado sería

$$\hat{y}(x_1 = -1, x_3 = -1) = \hat{\mu} - \frac{\hat{\alpha}}{2} - \frac{\hat{\gamma}}{2} + \frac{(\alpha\gamma)}{2}$$

En cuanto a la interpretación del gráfico:

- El hecho de que una recta sea superior a la otra indica que el factor C influye mucho
- El hecho de que los puntos de la derecha sean siempre superiores a los de la izquierda indica que el factor A también influye
- El hecho de que no sean paralelas las rectas indica que hay interacción entre ambas variables. Es decir, sólo se puede considerar que no hay interacción si ambas rectas fueran paralelas.

El modelo 2^k

El modelo, estimaciones, etc. son prácticamente iguales que el modelo anterior. Sólo hay algunas matizaciones que hacer. La primera es que en estos modelos las interacciones altas se suelen considerar nulas, en concreto las superiores a 3, lo que además supone que se dispone de observaciones para estimar el error experimental. Se suele utilizar el mismo estimador a partir de la MEDA para el error experimental salvo que en este caso será $\hat{\sigma} = \hat{s}_{\vartheta}\sqrt{2^{k-2}}$. Respecto a los contrastes sobre los parámetros, en la práctica se suponen significativos aquellos para los que $|\hat{\vartheta}_j| \geq 2\hat{s}_{\vartheta}$ y si k es grande se hace con un 3.

Una cuestión que resulta de interés es decidir si es mejor hacer un diseño 2^{k-1} replicado o introducir un factor más y hacer uno 2^k con el riesgo de añadir un factor *inerte*, es decir, que no ejerce influencia alguna. La respuesta es sencilla, es mejor hacer uno 2^k , ya que si el nuevo

facto es inerte ambos procedimientos son idénticos y tendremos el diseño replicado, y si es activo, bien por sí solo o por interacción el diseño añmplido permitirá medir su efecto.

Fracciones de diseños factoriales a dos niveles

El problema de los diseños anteriores es que requiere un número elevado de pruebas, especialmente si se utilizan muchos factores. La idea de los diseños fraccionales es realizar sólo una fracción del diseño completo que permita estimar los parámetros de orden bajo. En estos diseños se suponen nulas las interacciones de orden alto. Estos diseños se utilizan sobre todo en los principios de una investigación para identificar los factores que tienen mayor efecto en la respuesta.

Supongamos que en un diseño 2^3 seleccionamos como fracción aquellos experimentos en que ABC siempre tiene signo +. La tabla sería:

A	B	C	AB	AC	BC	ABC	Y
+	-	-	-	-	+	+	(a)
-	+	-	-	+	-	+	(b)
-	-	+	+	-	-	+	(c)
+	+	+	+	+	+	+	(abc)

Esta tabla requiere la mitad de observaciones. La estimación de los parámetros se realizaría como en el diseño completo, pero hay que tener en cuenta que ahora no se estima sólo el efecto de ese factor sino que incluirá también interacciones que no pueden separarse, es decir, ambos efectos están confundidos. Por ejemplo, al hacer la estimación del parámetro del factor A no es posible impedir que se incluya también el efecto de la interacción BC ya que tiene los mismos signos que A. Denotando por \hat{s}_A a la estimación hecha de la forma habitual pero dividiendo el resultado del algoritmo de los signos por 2 no por 4 (se cambia la notación porque no será la estimación directa del parámetro), los efectos confundidos para este diseño serán

$$\hat{s}_A \rightarrow A + BC \quad \hat{s}_B \rightarrow B + AC \quad \hat{s}_C \rightarrow C + AB \quad \hat{s}_\mu \rightarrow \mu + ABC$$

De haber la fracción de todos (-) en ABC los efectos confundidos serían

$$\hat{s}_A \rightarrow A - BC \quad \hat{s}_B \rightarrow B - AC \quad \hat{s}_C \rightarrow C - AB \quad \hat{s}_\mu \rightarrow \mu - ABC$$

Por lo tanto, al seleccionar una fracción sólo se podrán estimar los efectos principales si se consideran nulas las interacciones y los estimadores serán estos valores.

Para saber qué efectos son los confundidos se utiliza la *ecuación generatriz*. Así, en el caso anterior la ecuación generatriz será $I = ABC$ (I es una columna de signos +), y los efectos confundidos se obtienen multiplicando esta matriz por los factores con las reglas $AI = A$ y $AA = I$. Así se tiene $A = AI = AABC = IBC = BC$, $B = BI = BABC = AC$ y $C = IC = ABCC = ABI = AB$, es decir, la confusión será $A = BC$ $B = AC$ $C = AB$ $I = ABC$. Para diseños mayores se multiplica también por las interacciones. El procedimiento es análogo si se seleccionan los negativos pero se hablará de $-I$.

Es obvio que no resultará equivalente elegir una ecuación generatriz u otra ya que producirá distintos efectos confundidos. Para compararlas se utiliza el concepto de *resolución*. Se define la *resolución de una fracción factorial* como

Resolución = 1 + orden de interacción más baja confundida con algún efecto principal

Por ejemplo, la media fracción de un diseño $2^5 (2^{5-1})$ que se obtiene con la ecuación generatriz $I = ABCDE$ tiene resolución 5 (los efectos principales se confunden con las interacciones de cuarto orden); si fuera la media fracción obtenida con $I = ABCD$ sería de resolución 4. Obsérvese que la resolución coincide con el número de letras que hay en la ecuación generatriz (si es una palabra, si hay más, lo que no es posible en medias fracciones pero sí en cuartas, etc., será el número de letras que hay en la palabra más corta).

Obviamente, un diseño será mejor cuanto mayor resolución tenga, pues, confundirá los efectos principales con interacciones de mayor orden, y se supone que cuanto mayor sea el orden menor importancia tiene la interacción, e incluso se suele despreciar la interacción a partir de ciertos órdenes.

Construcción de diseños 2^{k-p}

Supóngase que se desean estudiar 5 factores con 8 experimentos. Esto supone que se desea un diseño del tipo 2^{5-2} . Hay dos procedimientos para determinar el diseño que se debe hacer: uno es elegir de un 2^5 qué dos interacciones poner todas positivas o negativas y otro es hacer un diseño 2^3 y estudiar cómo añadir las nuevas 2 variables. De ambos procedimientos es preferible el segundo. La razón fundamental es que en el primero es difícil determinar la resolución con la que se hace la selección. Por ejemplo, si se seleccionan $I = ABCDE$ junto con $I = ABCD$ al multiplicar por E en la segunda para ver los efectos confundidos se tendría $E = EI = EABCD = ABCDE = I$, con lo que la resolución sería 1 y no sería posible estimar el efecto de este factor.

Para obtener el diseño haciendo un 2^3 completo lo que hay que saber es cómo añadir las dos variables que faltan. Si se asigna una variable a la interacción de tercer orden ($D = ABC$) y otra a una de segundo orden (por ejemplo $E = AB$) la ecuación de definición de la fracción será $I = ABCD = ABE$, con lo que la resolución será 3. Obsérvese que de esta forma no es posible que no se pueda estimar el efecto principal de un factor, y proporciona directamente la ecuación generatriz que indica la resolución del diseño. Se llama *ecuación generatriz completa* a todas las posibles mutliplicaciones de las ecuaciones de selección de todas las formas posibles, de modo que en diseño 2^{6-3} definido por $D = AB \quad E = AC \quad F = BC$, la ecuación generatriz sería $I = ABD = ACE = BCF$ y la completa $I = ABD = ACE = BCF = BDCE = ACDF = ABEF = DEF$. El interés en conocer la ecuación generatriz completa radica en que para saber la estructura de confusión basta con multiplicar por la ecuación completa el efecto que se desea saber con quién está confundido. Por ejemplo, para saber con qué efectos se confunde A se multiplica por la ecuación completa, de modo que se obtiene $A = BD = CE = ABCF = ABDCE = CDF = BEF = ADEF$.

Análisis de fracciones factoriales

Para analizar qué factores o estimaciones resultan significativas se suele utilizar el método de la MEDA de modo que llamando $\hat{s}_\theta = MEDA/0.675$ se acepta como significativas aquellas estimaciones que en valor absoluto sean mayores que $2\hat{s}_\theta$, o $3\hat{s}_\theta$ si hay muchos datos.

Una vez decididas cuáles se consideran significativas hay que decidir de los efectos confundidos en ellas cuáles son los que se consideran significativos. Para ello se suelen seguir dos reglas: primera, que los efectos principales son mayores en magnitud que las interacciones y dentro de éstas las de órdenes más bajos mayores que las de órdenes más altos, y que es raro encontrar interacciones altas activas que sean combinaciones de efectos inactivos. Supongamos entonces el diseño 2^{3-1} con $I = ABC$. Los efectos confundidos en las estimaciones serán $\hat{s}_A \rightarrow A + BC \quad \hat{s}_B \rightarrow B + AC \quad \hat{s}_C \rightarrow C + AB \quad \hat{s}_\mu \rightarrow \mu + ABC$. Supongamos que sólo se considera activa \hat{s}_A , entonces caben tres interpretaciones: a) sólo A es activo, b) sólo la interacción BC es activa, c) ambas son activas. De las tres opciones se opta por la primera según las reglas enunciadas anteriormente.

Los diseños factoriales tienen múltiples aplicaciones en la industria. En ella básicamente hay dos formas de experimentación: fuera de la línea (off-line) y en la línea (on-line). La que se hace fuera de la línea suele llevarse a cabo por personal especializado, un gran número de variables y tiempo limitado, por lo que se suelen utilizar diseños fraccionales. La que se hace en

la línea suele hacerse por personal no especializado, ha de ser sencilla pues no puede influir en el tiempo de proceso, y suele ser sin limitación de tiempo, por lo que suelen usarse diseños simples (2^2 o 2^3) replicados varias veces, lo que se conoce como Operación Evolutiva o EVOP (por ejemplo para 2 factores se eligen 2 niveles para cada uno y se experimenta con un 2^2 , si pasado un tiempo se decide que el primer factor es significativo se sitúa en el nivel que mejor vaya y se selecciona otro nivel para él y se vuelve a repetir el experimento, así hasta que se de cómo no mejorable la configuración).

IV.7.9 Modelo de regresión

Estos modelos suponen que la variable respuesta es una función lineal de los factores de los que depende el modelo, factores continuos de modo que no pueden ser fijados a todos sus niveles y en muchos casos aleatorios, y de una perturbación. El modelo sería de la forma

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \text{ v.a.i.i.d. } N(0, \sigma^2)$$

El modelo de regresión pretende medir el efecto de las variable más importantes, determinando cuáles son éstas, y representa el efecto de las restantes (incluso desconocidas) en la perturbación.

Sólo se va a hacer una breve exposición de lo que es el modelo de regresión. Para una descripción detallada de las hipótesis del modelo, estimación de parámetros y diagnosis del modelo se debe consultar un libro de estadística, como Peña (1998).

Las preguntas que se pueden hacer en un modelo son ¿qué valores toman los coeficientes que ahí aparecen? ¿es este modelo válido?

Para responder a estas preguntas se estiman los parámetros y se hace un *análisis de la varianza*, con 2 factores en este caso, con el fin de determinar si los factores influyen en la respuesta. Después se analizan los residuos para ver si las hipótesis del modelo (normalidad de residuos, homcedasticidad, etc.) son válidas.

Sin embargo, en el modelo general de regresión existen además otros procedimientos para estimar los parámetros significativos del modelo, sobre todo para el caso en que las variables explicativas no sean independientes. En este caso hay que tener en cuenta que el efecto de una variable puede incluir el de otra con lo que hay que tener en cuenta que no se pueden estimar por separado y no se pueden duplicar efectos.

Para determinar el modelo hay básicamente dos procedimientos:

- a) estimar los parámetros del modelo conjuntamente y hacer un análisis de la varianza para determinar cuáles pueden considerarse significativos, prescindiendo entonces de los que no lo sean y estimando de nuevo los parámetros de las otras sin estos factores. Con este procedimiento la relación que pueda haber de dos variables conjuntamente aparece amortiguada en cada uno de los coeficientes de éstas.
- b) hacer una regresión por pasos, en la que en cada iteración se va añadiendo una variable al modelo (con la opción incluso de quitar alguna variable que se hubiera añadido antes si con la añadida y el resto pierde su significación), hasta que se considera que ninguna de las variables que no han sido incluidas en el modelo todavía pueden aportar información relevante (para determinar la relevancia de una variable no incluida se analiza la relación entre los residuos del modelo parcial con las variables restantes). Con este procedimiento, el coeficiente con que son incluidas las nuevas variables al modelo será sólo para explicar en los residuos lo que no hayan podido explicar las anteriores, de modo que lo que sería por correlación con otras variables quedaría en el coeficiente de las que ya están incluidas en el modelo.

Respecto al tamaño de la experimentación en regresión con múltiples factores, sólo decir que no es recomendable incluir en una regresión un número de variables de modo que el cociente entre el número de variables y el número de datos sea alto, ya que el coeficiente de determinación tiende a tomar un valor alto en esos casos y puede falsear la bondad del modelo (con pocos datos y muchas variables es fácil ajustar la función para que se aproxime a ellos).

V Construcción de modelos de simulación válidos y creíbles

El objetivo principal de la metodología de modelado es que un modelo debe ser una representación adecuada del sistema que se estudia. Sin embargo, una pregunta surge de forma inmediata, ¿cómo determinar cuándo un modelo de simulación es una representación suficientemente exacta del sistema, es decir, cuándo el modelo es válido?

No hay respuesta exacta a esta pregunta, sólo se pueden dar algunas pautas que seguir durante todo el proceso con el fin de hacerlo válido y creíble. Este capítulo se dedica precisamente a definir los conceptos asociados a validación, verificación y credibilidad, y dar algunas normas y consejos útiles a ser seguidos en el desarrollo de un modelo.

V.1 Definiciones

El desarrollo de un modelo de simulación pasa por distintas fases, pudiendo agruparse todas ellas en tres:

- *Fase de análisis y modelado*: en la que se obtiene un *modelo conceptual*, entendiendo por tal la representación matemática o lógica del problema, formulado para un estudio particular.
- *Fase de programación e implantación*: en la que se obtiene un modelo computacional, que es la traducción del modelo conceptual a un programa de ordenador e implantación en un ordenador.
- *Fase de experimentación*: en la que se hacen inferencias sobre el problema mediante experimentos computacionales.

Cada una de estas fases tiene asociado un concepto que representa el proceso de aceptación de la corrección de esa fase tanto por la persona que desarrolla el modelo como por el usuario o cliente del mismo. Estos conceptos son la verificación, la validación y la credibilidad.

La *verificación* consiste en determinar que *el programa de ordenador* se comporta como es debido, es decir, que se ha realizado una traducción correcta del modelo conceptual a un programa de ordenador que funciona correctamente.

La *validación* consiste en determinar si el *modelo conceptual* es una representación adecuada del sistema, es decir, si nuestro conocimiento del sistema se ha traducido en hipótesis que reproducen correctamente su comportamiento referente a los objetivos del estudio.

La *credibilidad* es hacer que los resultados sean aceptados por el decisor para ser utilizados en el proceso de toma de decisiones, es decir, han de ser creíbles.

La siguiente gráfica muestra las fases y los conceptos que intervienen en cada una de ellas.

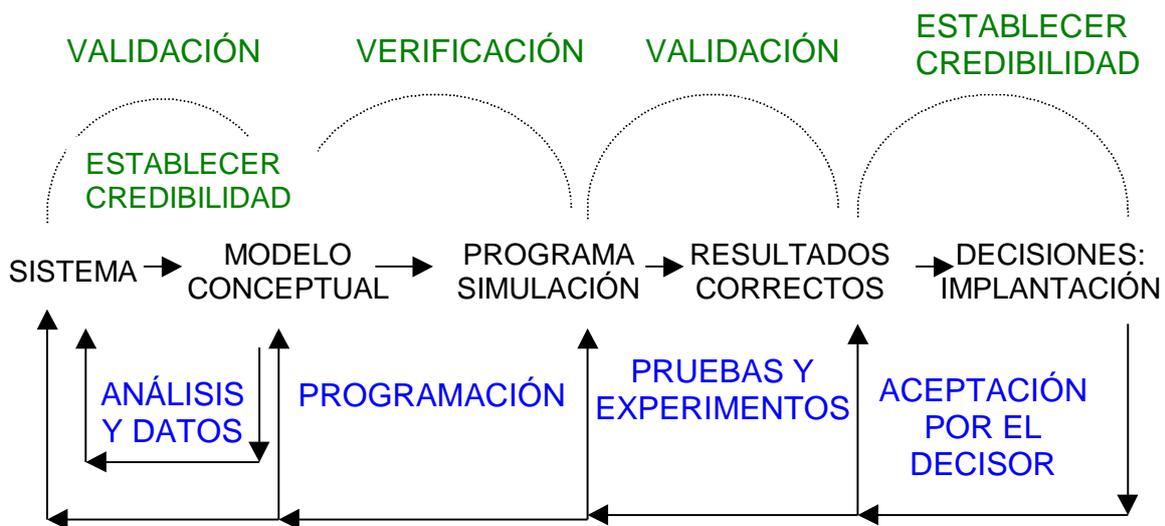


Figura 13. Fases y conceptos de aceptación asociados en el desarrollo de un modelo de simulación

Para poder validar, verificar y establecer la credibilidad hay una serie de pautas que se deben seguir en cada una de ellas, con el fin de facilitar esta labor.

Así pues, para la *validación* algunas pautas y consejos son los siguientes:

- En primer lugar, determinar si de todos los aspectos del sistema hemos incorporado los que son de real interés para nuestro estudio, sin dejarnos ninguno que pueda influir en el comportamiento; hay que tener en cuenta que se desea experimentar con el modelo como con un sustituto del sistema real.
- Los modelos se desarrollan con un objetivo, y es importante, por lo tanto, que haya una claridad de esos objetivos desde el principio, y que al definirlos se implique al responsable de la toma de decisiones, de ese modo se logrará también la credibilidad deseada.

- Por otra parte, ha de establecerse un equilibrio entre realismo (con la complejidad que implica) y objetivos, no incluyendo detalles innecesarios; no se debe perder nunca de vista que el modelo es siempre una aproximación.
- Por último, es recomendable utilizar en el desarrollo procedimientos jerárquicos, empezando con modelos sencillos para luego ir complicándolos.

En cuanto a la *verificación*, es un proceso más estándar y conocido que forma parte del mundo de la programación, pero algunas de las pautas habituales son las siguientes:

- Organizar el programa en submódulos, verificables por separado.
- Para asegurarse que está libre de errores de programación probarlo con casos sencillos, cuyo resultado sea conocido.
- Tener una traza de la ejecución para ver si sigue la lógica de los procesos; es habitual comparar en algún ejemplo sencillo la traza de la ejecución con una traza desarrollada a mano.
- Utilizar la animación gráfica para hacer la verificación mediante la visualización, y para dar credibilidad al simulador.

V.2 Metodología de validación y credibilidad

Aunque en la sección anterior se han dado algunas pautas generales acerca de la validación, pero éstas se pueden concretar más dando un procedimiento de validación que a su vez logre establecer la credibilidad buscada en el modelo. La metodología sería la siguiente:

1. Desarrollar desde el principio un modelo válido, es decir, con el objetivo de ser bueno, que sea creíble, es decir, razonable. Para ello se debe
 - involucrar a los usuarios
 - verificar la recogida de datos
 - si es posible, intentar basar aspectos del modelo en teorías bien establecidas
2. Verificar empíricamente las hipótesis en que se basa el modelo.
 - Verificar adecuadamente la bondad de ajuste de las distribuciones seleccionadas.
 - Realizar un análisis de sensibilidad, que permita identificar cuán sensible son los resultados de la simulación a algún aspecto del modelo, teniendo pues especial cuidado con los más relevantes; este análisis conlleva realizar un adecuado diseño de

experimentos. Por ejemplo, cuando hay dudas acerca de la distribución seleccionada, se deben realizar experimentos variando la distribución y comparar los resultados para determinar si el modelo realmente es sensible a una u otra distribución de la variable. En el caso en que sí lo sea, se deberá extremar el cuidado en la selección de la distribución, obteniendo si es posible nuevas observaciones que ayuden a tomar la decisión final.

3. Determinar cuán representativos son los resultados de la simulación: esto implica hacer una comparación entre resultados reales y resultados de simulación, según se explica a continuación.

V.2.1 Comparación entre datos reales y simulados

La comparación entre los valores reales y los valores obtenidos mediante simulación no es una tarea trivial, sino que debe hacerse con sumo cuidado para no llegar a conclusiones erróneas.

Existen fundamentalmente tres aproximaciones para hacer esta comparación:

1. Comparación por *inspección*:

Sea R_1, \dots, R_k el conjunto de datos reales, y π_1, \dots, π_k el conjunto de datos simulados.

Una primera idea sería aplicar los tests de homogeneidad estadísticos, que intentar establecer si dos muestras pueden provenir de la misma población. Sin embargo, esta aproximación no suele ser buena ya que no suele haber independencia entre los datos, que es una de las hipótesis de estos tests, la independencia entre ambas muestras.

Otra posibilidad fácil de llevar a cabo es hacer comparaciones directas de medias, varianzas, etc. Sin embargo, es muy peligroso usar este sistema ya que es muy vulnerable a la aleatoriedad.

2. *Inspección correlacionada*:

Consiste en establecer las comparaciones alimentando el modelo con los datos históricos. El esquema sería el siguiente

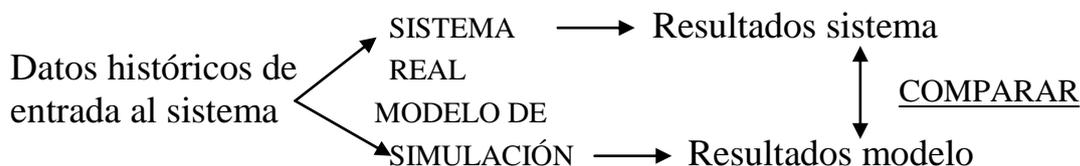


Figura 15. Inspección correlacionada de datos reales y datos simulados

3. *Análisis de intervalos de confianza basados en datos independientes:*

Sean $\{R_1, \dots, R_m\}$ m conjuntos independientes del sistema, y $\{M_1, \dots, M_n\}$ n conjuntos independientes del modelo. Considérese la esperanza de cada uno de los $m + n$ conjuntos, $\mu_R = E[R_j]$ y $\mu_M = E[M_j]$.

El objetivo de este análisis es obtener un intervalo de confianza para la diferencia de ambas medias, $\xi = \mu_R - \mu_M$.

El intervalo de confianza a nivel $1 - \alpha$ será

$$\bar{R} - \bar{M} \pm t_{f, \alpha/2} \sqrt{\frac{S_R^2}{m} + \frac{S_M^2}{n}}$$

donde el número de grados de libertad de la t de Student viene dado por la expresión

$$f(x) = \begin{cases} \frac{x}{8} & x \in [0, 2] \\ \frac{1}{2} & x \in (2, 8] \\ \frac{5}{4} - \frac{x}{8} & x \in (8, 10] \\ 0 & \text{resto} \end{cases} .$$

4. *Análisis de series cronológicas:*

Este análisis se basa en un análisis espectral que excede los objetivos de este capítulo

VI Aplicaciones de la simulación en finanzas¹⁰

VI.1 Nociones básicas sobre derivados financieros

En las actividades financieras hay múltiples mecanismos de inversión, la mayoría de ellos denominados *derivados financieros* puesto que su valor depende o se deriva del valor de otro activo financiero más básico denominado *activo subyacente*. El subyacente utilizado puede ser muy diferente: acciones, índices bursátiles, valores de renta fija, tipos de interés o también materias primas (oro, petróleo, productos agrícolas...).

Las características generales de los derivados financieros son:

- Su valor cambia en respuesta a los cambios de precio del activo subyacente.
- Requiere una inversión inicial neta muy pequeña o nula, respecto a otro tipo de contratos que tienen una respuesta similar ante cambios en las condiciones del mercado.
- Se liquidará en una fecha futura.
- Pueden cotizarse en mercados organizados (como las bolsas) o no organizados ("OTC")

Los principales derivados son las siguientes:

- *Contrato de futuros (forward)*: contrato o acuerdo que obliga a las partes contratantes a comprar o vender un número determinado de bienes o valores (*activo subyacente*) en una fecha futura y determinada y con un precio establecido de antemano.
- *Opciones*: contrato que da a su comprador el derecho, pero no la obligación, a comprar o vender una cantidad determinada del *activo subyacente* a un precio predeterminado (*strike* o *precio de ejercicio*), en o hasta una fecha concreta (*vencimiento* o *fecha de ejercicio*).
- *Intercambios (swaps)*: contrato por el cual dos partes se comprometen a intercambiar una serie de cantidades de dinero en fechas futuras. Normalmente los intercambios de dinero futuros están referenciados a tipos de interés, llamándose **IRS** (Interest Rate Swap)

¹⁰ Basado en Camaño (2006)

aunque de forma más genérica se puede considerar un **swap** cualquier intercambio futuro de bienes o servicios (entre ellos el dinero) referenciado a cualquier variable observable.

VI.1.1 Contratos de futuros

Un **contrato de futuros** es un contrato o acuerdo que obliga a las partes contratantes a comprar o vender un número determinado de bienes o valores (*activo subyacente*) en una fecha futura y determinada y con un precio establecido de antemano.

Existen dos motivos por los cuales alguien puede estar interesado en contratar un futuro:

- Operaciones de cobertura: La persona tiene o va a tener el bien subyacente en el futuro (petróleo, gas, naranjas, trigo, etc.) y lo venderá en un futuro. Con la operación quiere asegurar un precio fijo hoy para la operación de mañana.
- Operaciones especulativas: La persona que contrata el futuro sólo busca especular con la evolución de su precio desde la fecha de la contratación hasta el vencimiento.

Dentro del argot financiero, es importante entender lo que se llaman posiciones largas y cortas.

- **Estar largo.** Quien compra contratos de futuros, adopta una posición larga, por lo que tiene el derecho a recibir en la fecha de vencimiento del contrato el activo subyacente objeto de la negociación. Básicamente significa comprar hoy para vender mañana o invertir hoy para mañana recuperar el nominal más las plusvalías.
- **Estar corto.** Quien vende contratos adopta una posición corta ante el mercado, por lo que al llegar la fecha de vencimiento del contrato deberá entregar el correspondiente activo subyacente, recibiendo a cambio la cantidad de dinero acordada en la fecha de negociación del contrato de futuros. Básicamente significa financiarse hoy con la venta del activo que aún no tenemos, tomando la obligación de devolver el activo mañana.

Con independencia de que un contrato de futuros se puede comprar con la intención de mantener el compromiso hasta la fecha de su vencimiento, también puede ser utilizado como instrumento de cobertura en operaciones de tipo especulativo, ya que no es necesario mantener la posición abierta hasta la fecha de vencimiento; en cualquier momento se puede cerrar la posición con una operación de signo contrario a la inicialmente efectuada: cuando se tiene una posición compradora, puede cerrarse la misma sin esperar a la fecha de vencimiento simplemente vendiendo el número de contratos compradores que se posean; de forma inversa, alguien con una posición vendedora puede cerrarla anticipadamente acudiendo al mercado y comprando el número de contratos de futuros precisos para compensar su posición.

VI.1.2 Opciones

Una opción es un contrato que da a su comprador el derecho, pero no la obligación, a comprar o vender una cantidad determinada del *activo subyacente* a un precio predeterminado (*strike* o *precio de ejercicio*), en o hasta una fecha concreta (*vencimiento*).

Una opción se compra (o vende) a un precio o *prima*, que es el valor inicial de la opción. El precio del subyacente es observable en el tiempo y va variando siguiendo un proceso estocástico. Según varía el precio del subyacente va variando el valor de la opción.

Las opciones se pueden clasificar de varias formas. De forma muy genérica se pueden dividir en dos grupos: las "vainilla" que consisten en los contratos básicos de opciones "call" o "put" y las "exóticas" que incorporan variantes que hacen más complejo su tratamiento y su valoración.

VI.1.2.1 Opciones "vainilla" (*vanilla options*)

Se trata de las opciones "básicas" que, dependiendo del tipo de derecho que nos den, son *opciones call* (de compra) y *opciones put* (de venta). En función de su forma de ser ejercidas podemos diferenciar:

- **Opción europea:** tan sólo se pueden ser ejercidas en una fecha determinada (fecha de ejercicio).
- **Opción americana:** pueden ser ejercidas a lo largo de su vida hasta la fecha de ejercicio.

OPCIÓN CALL

Una opción call da a su comprador el *derecho* -pero no la obligación- a comprar un activo subyacente a un precio predeterminado en una fecha concreta. El vendedor de la opción call tiene la *obligación* de vender el activo en el caso de que el comprador ejerza el derecho a comprar.

Se llama *pay-off* a la ganancia al vencimiento de la opción (sin incluir el coste de la prima). Por ejemplo, si K es el precio de ejercicio y S el valor del activo subyacente en el momento del ejercicio, el *pay-off* de una opción call será $\max(S - K, 0) = (S - K)^+$

La compra de una opción call es interesante cuando se tienen expectativas alcistas sobre la evolución futura del mercado de valores.

Posibles situaciones favorables para la compra de opciones call:

- Cuando se prevé que una acción va a tener una tendencia alcista, ya que es más barato y rentable que la compra de acciones.
- Cuando una acción ha tenido una tendencia alcista fuerte, el inversor no ha comprado y puede pensar que está cara, pero que puede seguir subiendo, la compra de una call permite aprovechar las subidas si la acción sigue subiendo y limitar las pérdidas si la acción cae.
- Cuando se quiere comprar acciones en un futuro próximo porque se cree que van a subir pero hoy NO se dispone de los fondos necesarios, la opción call permite aprovechar las subidas sin tener que comprar las acciones.

La compra de una opción call implica:

a) Se puede comprar la acción a un precio fijo. Este precio (*precio de ejercicio*) lo fija el comprador.

b) Todo lo que la acción suba en la Bolsa por encima de dicho precio de ejercicio menos el precio pagado por la prima son ganancias.

c) Si el precio de la acción cae por debajo del precio de ejercicio, las pérdidas son limitadas y conocidas: son exactamente igual al precio pagado por la opción, es decir, la *prima*, ya que no se ejercería la opción.

d) El coste de la opción es mucho menor que el de la compra de la acción.

e) El apalancamiento (relación coste de la inversión/rendimiento) es muy alto. Con pequeñas inversiones pueden obtenerse altas rentabilidades.

En la venta de una opción call, el vendedor recibe la prima (el precio de la opción). A cambio, está obligado a vender la acción al precio fijado (precio de ejercicio), en el caso de que el comprador de la opción call ejerza su opción de compra.

Posibles situaciones favorables para la venta de opciones call:

- Para asegurar ingresos adicionales una vez decidida la venta de las acciones.
- Es el caso de que no importe vender las acciones a un precio considerado suficientemente alto y recibir, además, un ingreso extra previo. Este es el caso en que se vende una call fijando un precio de ejercicio en el nivel que se desee por encima del precio actual de la acción en Bolsa. Si la acción llega a alcanzar ese precio, habrá que vender la acción, pero a un precio alto y, además, se habrá ingresado el valor de la opción.

La venta de una opción call supone:

- Genera un flujo monetario inmediato derivado del ingreso procedente de la venta de la opción.
- Retrasa el momento en que se entra en pérdidas por bajadas en el precio de la acción.
- Proporciona una atractiva rentabilidad si la acción se mantiene estable.

OPCIÓN PUT

Una opción put da a su comprador el *derecho* -pero no la obligación- a vender un activo a un precio predeterminado hasta una fecha concreta. El vendedor de la opción put tiene la *obligación* de comprar el activo en el caso de que el comprador de la opción decida ejercer el derecho a vender el activo.

El *pay-off* o ganancia al vencimiento de la opción (sin incluir el coste de la prima), si K es el precio de ejercicio y S el valor del activo subyacente en el momento del ejercicio, será $\max(K - S, 0) = (S - K)^-$

La compra de opciones put se utiliza como *cobertura*, cuando se preven caídas de precios en acciones que se poseen, ya que mediante la compra de Put se fija el precio a partir del cual se gana dinero. Si la acción cae por debajo de ese precio, el inversor gana dinero. Las pérdidas quedan limitadas a la prima. Las ganancias aumentan a medida que el precio de la acción baja en el mercado.

Por tanto, es interesante comprar una opción put:

- Cuando se tiene acciones y se cree que hay grandes probabilidades de que su precio caiga a corto plazo, pero se piensa que el valor tiene una tendencia alcista a largo plazo, por lo que no se quiere vender dichas acciones. Con la opción put se obtienen beneficios si caen los precios y no se tiene que vender las acciones. De este modo se aprovecharía la futura subida de los precios de la acción. Es una forma de proteger beneficios no realizados cuando se tienen acciones compradas. A esta operación se le conoce como "Put protectora", porque protege la inversión de caídas.
- Cuando se está convencido de que la acción va a caer y se quiere aprovechar esa caída para obtener beneficios. Si no se tienen acciones compradas previamente también interesa comprar una opción put, pues con ello se obtienen beneficios con las caídas de la acción.

El vendedor de una opción put está vendiendo un derecho por el que cobra la prima. Puesto que vende el derecho, contrae la obligación de comprar la acción en el caso de que el comprador de la put ejerza su derecho a vender.

Posibles situaciones favorables para la venta de opciones put:

- Para comprar acciones con descuento. Cuando interese comprar acciones a un precio fijo por debajo del nivel actual de precios y además con un descuento. El descuento es la prima ingresada por la venta de la opción.
- Cuando se piensa que el precio de la acción va a entrar en un período de estabilidad, se está convencido de que no va a caer y que es posible que tenga ligeras subidas. En esta situación se puede fijar un precio a partir del cual se está dispuesto a comprar; entretanto, se ingresa la prima. El precio límite de compra es el precio de ejercicio al que se venderá la opción put.

VI.1.2.2 Opciones “exóticas” (*exotic options*)

Son opciones que son más complejas que las opciones comúnmente negociadas (*plain vanilla*). Estos productos son negociados normalmente *over-the-counter* (OTC). Incorporan distintas variantes ("exoticidades") que pueden llegar a complicar el cálculo de la valoración de la opción en gran medida.

- Exoticidad en el cálculo del pago (*pay off*):
 - Opción “asiática” (*Asian option*): depende de la media del valor del subyacente en un periodo determinado.
 - *Lookback option*: se calcula en función del máximo (o mínimo) alcanzado por el subyacente en un periodo. Variantes: opción rusa (*Russian option*) es una opción lookback que está operativa en perpetuidad.
 - Opción binaria o digital (*digital / binary option*): el pago puede ser una cantidad determinada (o un activo) o, por el contrario, no haber pago en absoluto.
 - Opción “oscilante” (*swing option*): el comprador puede “oscilar” el precio del subyacente. Principalmente empleada en energía.
 - Opción “parisina” (*Parisian option*): depende del tiempo que el activo esté por encima (o por debajo) del *strike*.

- Exoticidad en la fecha/forma de ejercicio:
 - Opción “bermuda” (**Bermudan option**): permite ser ejercida en varios momentos del tiempo (espaciados de forma discreta); por ejemplo, trimestralmente. Variantes: opción “canaria” (**Canary option**) es una opción a caballo entre una opción europea clásica y una bermuda; permite ser ejercida en varios momentos pero nunca antes de un periodo fijo, por ejemplo, de un año.
 - Opción con barrera (**barrier option**): la opción deja de existir –*knock out*- (o comienza a existir –*knock in*-) cuando el subyacente alcanza (o se cruza) un determinado valor (*barrier level*). Se pueden dar distintas combinaciones de condiciones:
 - **Up-and-out**: el subyacente comienza a fluctuar bajo el *barrier level* y si lo alcanza, la opción deja de existir (*knock out*).
 - **Down-and-out**: el subyacente comienza a fluctuar sobre el *barrier level* y si lo cruza, la opción deja de existir (*knock out*).
 - **Up-and-in**: el subyacente comienza a fluctuar bajo el *barrier level* y si lo alcanza, la opción se activa (*knock in*).
 - **Down-and-in**: el subyacente comienza a fluctuar sobre el *barrier level* y si lo cruza, la opción se activa (*knock in*).
 - Opción “de estilo cap” (**capped-style option**): la opción se ejecuta automáticamente cuando el subyacente alcanza un determinado precio y se marca un nuevo *mark to market*. No tiene nada que ver, a pesar de su nombre, con un *interest rate cap*.
 - Opción “compuesta” (**compound option**): consiste en una opción sobre otra opción, suponiendo, de este modo, dos fechas de ejercicio y modalidades distintas.
 - Opción “grito” (**shout option**): Consiste en dos fechas de ejercicio distintas. El comprador puede señalar –o “gritar”- una fecha en la que el precio del subyacente le parezca interesante. En el momento final de madurez de la opción, el comprador puede decidir si le conviene el pago (*pay off*) a precio de la fecha final o a precio de la fecha del “grito”.
- Exoticidad en función del subyacente:
 - Opción “cesta” (**basket option**): se basa en una media ponderada de distintos subyacentes. Variantes: opción “arco iris” (**rainbow option**) se basa en una cesta en la que la ponderación de los componentes dependen de su comportamiento final. Por

ejemplo, un caso especial de este tipo de opción que es bastante común es una opción basada en el subyacente que peor comportamiento haya tenido dentro de una cesta. Otras variantes: *Himalayan option*, *mountain range option* ...

- Opción “de intercambio” (*exchange option*): se intercambia un activo por otro.

- Exoticidad en función de la divisa:
 - Opción “cruzada” (*cross / composite option*): el subyacente se negocia en una divisa y el *strike* está denominado en otra. Variante: opción “quanto” (*quanto option*) es una opción en el que el tipo de cambio queda fijado desde el comienzo; por ejemplo, 1.

- Otras exoticidades:
 - *Game / Israeli option*: el comprador puede cancelar la opción, afectando esto al *pay off* o teniendo que pagar algún tipo de "multa".
 - *Reoption*: es la posibilidad de renovar una opción que expiró sin haber sido ejercida.
 - *Chooser option*: da la opción de decidir si la opción será put o call.
 - *Forward starting option*: el *strike* se decide en el futuro, no en el comienzo del contrato. Variantes: *cliquet option* es una secuencia de *forward starting options*.

VI.2 Valoración de opciones y futuros

VI.2.1 Conceptos básicos en valoración de derivados

Una opción se compra (o vende) a un precio o *prima*, que es el valor inicial de la opción. El precio del subyacente es observable en el tiempo y va variando siguiendo un proceso estocástico. Según varía el precio del subyacente va variando el valor de la opción.

El cobro a que da lugar una opción en el momento del ejercicio por parte del comprador se denomina “*pay-off*”. Obviamente, el día del vencimiento de la opción el “*pay-off*” es igual al valor de la opción. En una opción europea se ejercerá el derecho de la opción si el *pay-off* o valor en la fecha de vencimiento es positivo.

En una opción americana o bermudas el ejercicio de la opción puede llevarse a cabo antes de la fecha de vencimiento. En tal caso, en cada momento que se puede ejercitar hay un *pay-off* que viene determinado por el valor del subyacente. Aunque puede darse el caso de que incluso con un *pay-off* positivo no se ejerza la opción con el fin de ejercerla más adelante. La curva de

precios del subyacente a lo largo del tiempo determina los valores para los que es óptimo ejercer la acción es lo que se llama **frontera de ejercicio óptimo**. Este precio marca la frontera entre la conveniencia o no del ejercicio en ese momento. En la valoración de opciones americanas y bermudas no se busca sólo su valor sino también la frontera de ejercicio óptimo. Los problemas para este tipo de opciones se conocen como problemas de frontera libre y son mucho más complejos que para las opciones europeas.

Como tratamos con cantidades en distinto momento del tiempo, hay que tener en cuenta que éstas han de ser actualizadas con la tasa de interés libre de riesgo, que llamaremos r . En general, este interés r vendrá dado como interés anual. Es decir, una cantidad K al vencimiento de una opción, en t años antes de esa fecha tiene un valor de $\frac{K}{(1+r)^t}$.

Sin embargo, en muchas ocasiones se necesita el interés en un tiempo continuo, es decir, con una expresión válida para cualquier incremento de tiempo t no necesariamente en años. Para ello se utiliza lo que se llama el interés en continuo, cuya relación con el interés anual es $e^{r_c t} = (1+r)^t$, o lo que es lo mismo haciendo $t=1$, $r_c = \ln(1+r)$. En general, en todo lo que se refiere a diferenciales y variaciones instantáneas utilizaremos la notación r pero refiriéndonos al interés en continuo.

Igualmente, para el caso en que se quiera obtener el interés o rendimiento en un intervalo Δt , por ejemplo, para un mes sería $\Delta t = 1/12$ entendiendo el año como unidad de tiempo, el rendimiento en ese periodo visto como interés tiene que cumplir $1 + r_{\Delta t} = e^{r_c \Delta t} = (1+r)^{\Delta t}$.

Gran parte de la teoría financiera actual descansa en la Hipótesis del Mercado Eficiente (“Efficient Market hypothesis” EMH) que viene a decir:

“En un mercado con inversores que actúan racionalmente, los precios de las acciones reflejan en todo momento la información disponible y se dice que el mercado es eficiente. En un mercado eficiente ningún tipo de análisis puede conducir a estrategias que batan consistentemente a un índice apropiado (“benchmark”)”

Los participantes compran bajo la creencia de que lo que adquieren vale más de lo que están pagando por ello y venden convencidos de que reciben más de lo vale, pero si los mercados son eficientes y los precios reflejan toda la información disponible, **compradores y vendedores intentando beneficiarse convierten al mercado en un juego de suerte y no de habilidad**.

Si un mercado es eficiente, la trayectoria seguida a lo largo del tiempo por los precios de los valores debe seguir un proceso estocástico en el que la memoria del pasado no exista. Este tipo de procesos se conocen como **procesos markovianos**.

De la EMH no se deduce que las variaciones absolutas de los precios o los rendimientos relativos en un período tengan que seguir una distribución normal, sin embargo, el modelo más extendido sobre los precios bursátiles, aunque muy discutido, es que los precios bursátiles siguen un tipo de proceso de difusión conocido como movimiento geométrico browniano (GBM) cuya ecuación diferencial estocástica viene dada por:

$$dS = \mu S dt + \sigma S dW \text{ o en términos de variación relativa } \frac{dS}{S} = \mu dt + \sigma dW \text{ siendo } \mu \text{ la}$$

tendencia o “drift”, σ la volatilidad y W un proceso de Wiener (proceso markoviano cuyos incrementos tienen distribución normal de media 0 y varianza el tamaño del incremento de tiempo). Esta expresión viene a decir que la variación relativa del activo sigue una distribución $N(\mu dt, \sigma \sqrt{dt})$, y las variaciones absolutas de éste $N(S\mu dt, S\sigma \sqrt{dt})$

Sin embargo, estas variaciones son en periodos estancos, si se desea saber el rendimiento acumulado $R(t)$ en un intervalo $[0, t]$, se tiene

$$S(t) = S(0)e^{R(t)} \text{ o equivalentemente } R(t) = \ln\left(\frac{S(t)}{S(0)}\right) = \ln(S(t)) - \ln(S(0)).$$

Aplicando resultados de cálculo estocástico que exceden el contenido de este capítulo (en concreto el Lema de Ito del Cálculo Estocástico aplicable sólo si se trabaja con un proceso de Wiener), se llega a la ecuación diferencial estocástica que expresa las variaciones del

$$\text{rendimiento compuesto } dR = \left(\mu - \frac{1}{2}\sigma^2\right) dt - \sigma dW.$$

Integrando el rendimiento acumulado se tiene

$$R(t) = \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) dt + \int_0^t \sigma dW = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma(W(t) - W(0)),$$

por lo que retomando el valor del activo se tiene

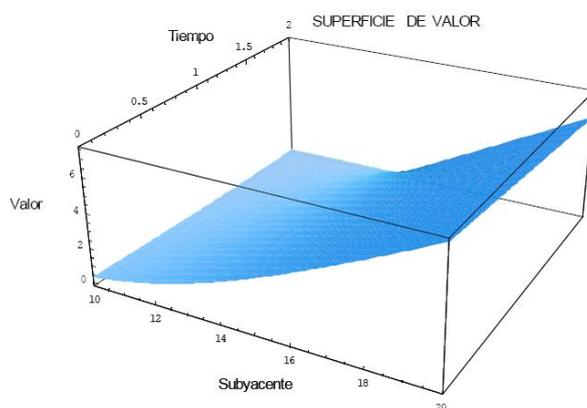
$$S(t) = S(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma(W(t) - W(0))}$$

Dadas las propiedades de los incrementos en un proceso de Wiener, el valor del activo sigue una distribución log-normal¹¹. Por otra parte, dado que $W(0) = 0$ c.s., la expresión se reduce a

$$S(t) = S(0)e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)}$$

Otra hipótesis asumida en la valoración de opciones es la ausencia de arbitraje. Un *arbitraje* es una estrategia de operaciones financieras, sin riesgo de crédito ni riesgo de mercado, con una probabilidad nula de dar una rentabilidad inferior a la tasa libre de riesgo y una probabilidad positiva de dar una rentabilidad superior a la tasa libre de riesgo correspondiente. Un mercado donde hay posibilidad de arbitraje se dice que no está en equilibrio, ya que si se da esa posibilidad, los arbitrajistas intervendrían modificando precios hasta restablecer el equilibrio. Por lo tanto, para la valoración de opciones se va a suponer la ausencia de arbitraje.

A la hora de valorar un contrato se busca una función de valor $V(S, t, T, K, \mu, \sigma, r)$, pero dado que todos son parámetros del contrato excepto el precio del subyacente y el tiempo, se considera una función $V(S, t)$ como la función de valor de una opción. El siguiente gráfico (Camaño (2006)) muestra una función de valor de una opción call a dos años con strike 14. Como puede observarse, al vencimiento el valor de la opción es el pay-off.



Para opciones europeas sobre subyacentes que no dan dividendos existe solución analítica, conocida como ecuación de Black&Scholes para la función de valor. Esta expresión es:

¹¹ Obsérvese que la esperanza de una lognormal es la exponencial de la media de la normal asociada más la mitad de la varianza de esa normal, por lo que se tiene el siguiente resultado, por otra parte, evidente $E[S(t)] = S(0) \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \frac{1}{2}\sigma^2(\sqrt{t})^2\right] = S(0)e^{\mu t}$

$$V(S, t) = \frac{e^{-rt}}{\sigma\sqrt{2\pi t}} \int_0^{\infty} e^{-\frac{\left(\ln\left(\frac{S}{X}\right) + \left(r - \frac{1}{2}\sigma^2\right)t\right)^2}{2\sigma^2 t}} \frac{1}{X} \text{Payoff}(X) dX$$

Así, en una opción call europea $\text{Payoff}(X) = \max(X - K, 0) = (X - K)I_{(K, \infty)}(X)$, y sustituyendo y operando en la expresión anterior se llega a $V_{\text{call}}(S, t) = S\Phi(d_1) - Ke^{-rt}\Phi(d_2)$,

siendo Φ la función de distribución de una $N(0,1)$, $d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}$ y

$$d_2 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}} = d_1 - \sigma\sqrt{t}$$

Para una opción put se puede repetir el proceso o utilizar la relación de paridad existente entre ambos tipos de opciones.

En la valoración de opciones europeas, se da la denominada *relación de paridad*. Si se tienen dos opciones con el mismo strike K , mismo vencimiento, pero una es call comprado y otra put vendido ambas sobre el mismo subyacente, obsérvese que si S es el precio del suyacente, el pay-off de la cartera es $C - P = \max(S - K, 0) - \max(K - S, 0) = S - K$. Así en cualquier momento anterior, y en particular al inicio, se tiene que cumplir la relación de paridad

$C - P = S - \frac{K}{(1+r)^t} = S - e^{-rt}K$. Esta relación no tiene por qué darse en opciones americanas o bermudas.

Por lo tanto, aplicando que $V_{\text{call}}(S, t) - V_{\text{put}}(S, t) = S - Ke^{-rt}$, y operando se llega a $V_{\text{put}}(S, t) = -S\Phi(-d_1) + Ke^{-rt}\Phi(-d_2)$

Sin embargo, el objetivo en este capítulo no es derivar o utilizar esta expresión analítica, sino estimar el valor de una opción mediante la simulación del activo subyacente y la estimación del pay-off obtenido.

Para ello todavía tenemos que dar algún concepto más del mercado como es el teorema fundamental de las finanzas y la medida de probabilidad de riesgo neutro. La idea básica es que **el valor de una opción puede computarse como el valor presente a la tasa libre de riesgo del valor esperado del "pay-off" computado con un drift igual a la tasa libre de riesgo**. Es decir,

$$V_{\text{opción}} = e^{-rt} E_{\text{riesgo neutro}}(\text{Payoff}(S))$$

No se trata de que las acciones sigan trayectorias GBM con “drift” igual a la tasa libre de riesgo, no es así. Lo que sucede es que la valoración puede realizarse suponiendo riesgo neutro y el resultado de la valoración es correcto.

Calcular el valor esperado del “pay-off” con estas trayectorias es calcular el valor esperado de un funcional de las trayectorias (“pay-off”) utilizando una medida de probabilidad nueva que denominamos **medida de probabilidad de riesgo neutro**. Una medida de probabilidad Q se dice que es de riesgo neutro si el precio libre de arbitraje de un derivado es el valor presente del valor esperado bajo Q del “pay-off” futuro del derivado.

En esta medida de probabilidad todas las acciones siguen GBM con “drift” igual a la tasa libre de riesgo (los valores presentes a la tasa de interés libre de riesgo del valor de una acción siguen en esta medida de probabilidad de riesgo neutro un proceso con “drift” nulo, es decir, una martingala).

Todos los activos bajo Q tienen la misma tasa esperada de retorno igual a la tasa libre de riesgo con independencia del nivel de riesgo asociado al mismo. Por este motivo se denomina a la medida Q medida de probabilidad de riesgo neutro.

El **teorema fundamental de la valoración de activos** se enuncia: Dado un mercado financiero se tiene que:

1. No existen arbitrajes si y solo si existe una medida de probabilidad de riesgo neutro equivalente a la medida de probabilidad real, es decir tal que los sucesos de probabilidad nula de una y otra medida son los mismos.

2. La medida de probabilidad de riesgo neutro es única si y solo si existen carteras de réplica de todos los productos derivados que permiten su cobertura, es decir si y solo si el mercado es completo.

De estos resultados se deduce que una valoración se puede obtener como $V_{opción} = e^{-rt} E_{riesgoneutro}(Payoff(S))$ donde S se rige por las ecuaciones diferenciales

$dS = rSdt + \sigma SdW$ o en términos relativos $\frac{dS}{S} = rdt + \sigma dW$. Y asumiendo que W fuera un

proceso de Wiener $S(t) = S(0)e^{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma(W(t) - W(0))}$.

Por último, dentro de estas nociones básicas, si el activo produce unos dividendos (o un interés si fuera una divisa) del tipo d como proporción del valor del activo, la posesión de éste puede verse como un interés adicional o lo que es lo mismo se puede ver como tener un “drift” de $r - d$. Si se tratara de una mercancía que por el contrario tuviera un coste c su posesión (coste de mantenimiento, etc.) el “drift” que habría de considerarse sería $r + c$. Estas

consideraciones son válidas para la estimación del payoff, obviamente no para la actualización del precio, es decir, han de utilizarse para las trayectorias, no para la actualización del valor de la opción.

El modelo de Black-Scholes supone la resolución de la esperanza planteada anteriormente bajo las hipótesis de que el mercado es eficiente y que el subyacente evoluciona según el proceso de Wiener, para una opción europea vainilla el valor de la opción se puede obtener

como $V_{call}(S, t) = Se^{-dt}\Phi(d_1) - Ke^{-rt}\Phi(d_2)$, siendo $d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - d + \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}$ y

$d_2 = d_1 - \sigma\sqrt{t}$, donde Φ hace referencia a la función de distribución de una normal estándar.

VI.2.2 Estimación del valor de opciones sin posibilidad de ejercicio anticipado mediante simulación

Se asume que el mercado financiero considerado admite una medida de probabilidad de riesgo neutro equivalente a la medida de probabilidad real, es decir, que ambas tienen el mismo conjunto de sucesos de probabilidad nula. En estas circunstancias, el teorema fundamental de la valoración de activos garantiza que el valor, en ausencia de arbitrajes, de cualquier contrato contingente viene dado por el valor presente a las tasa de interés libres de riesgo del valor esperado del “pay-off” en la medida de probabilidad de riesgo neutro. Es decir, que

$V_{opción} = e^{-rt} E_{riesgo\ neutro}(Payoff(S))$ siendo la ecuación diferencial estocástica de la variación

de S de la forma $\frac{dS}{S} = rdt + \sigma dW$.

No se trata de que el subyacente siga en la realidad este proceso sino de que se puede seguir el siguiente algoritmo para estimar mediante simulación el valor de una opción:

1. Simular la variable aleatoria “pay-off” a partir de sucesivas simulaciones del valor del subyacente al vencimiento (asumiendo un drift igual a la tasa libre de riesgo) o de la trayectoria discretizada del mismo según sea requerido para el cálculo del “pay-off”.
2. Estimar el valor esperado del “pay-off” mediante la media de la muestral de “pay-off” correspondientes a las simulaciones.

3. Computar el valor presente a la tasa libre de riesgo de dicho valor esperado. Obsérvese que esta actualización se puede hacer como $V_{opción} = e^{-r_c t} \overline{Payoff(S)}$ o como

$$V_{opción} = (1+r)^{-t} \overline{Payoff(S)}$$

Este cálculo arroja una estimación del valor del contrato tanto mejor cuanto mejor sea la estimación del valor esperado del “pay-off”. No importa lo complicado que sea el funcional que se utilice para el cálculo del “pay-off”, puede ser el simple de una opción “call” europea o el promedio aritmético de los valores de la trayectoria que es el “pay off” de una opción asiática, la valoración es siempre el valor presente a la tasa libre de riesgo del valor esperado con probabilidad riesgo neutro de dicho funcional.

En casos de contratos de tipo europeo en los que el “pay-off” dependa sólo del valor final no es necesario simular toda la trayectoria seguida por el subyacente durante la vida del contrato, sería suficiente simular el valor del subyacente en el día del vencimiento, si se conoce su distribución. Por ejemplo, si se trata de un subyacente que sigue un GBM en la medida de probabilidad de riesgo neutro gobernado por la ecuación diferencial estocástica

$$\frac{dS}{S} = r_c dt + \sigma dW \text{ donde } r \text{ es la tasa de interés libre de riesgo en continuo se sabe que su}$$

$$\text{solución viene dada por: } S(t) = S(0)e^{\left(r_c - \frac{1}{2}\sigma^2\right)t + \sigma(W(t) - W(0))}$$

$S(t)$ es una variable aleatoria que puede simularse directamente puesto que $W(t)$ es una variable aleatoria normal cuya varianza es t , y por lo tanto, no es necesario simular toda la trayectoria para simular el valor final del subyacente.

En otros casos de contratos también de tipo europeo, es decir, sin la posibilidad de ejercicio anticipado, el “pay-off” puede depender de la trayectoria seguida por el subyacente durante la vida del contrato, en este caso no basta con simular el valor final hay que simular la trayectoria.

También puede ocurrir que la ecuación diferencial estocástica que gobierna el proceso seguido por el subyacente en la medida de probabilidad de riesgo neutro no sea tan sencilla como un GBM cuya solución es conocida y no se disponga de la forma explícita de la distribución de $S(t)$. De hecho, lo normal en una difusión de Ito es que se desconozca la solución explícita. Por ejemplo, en el caso que el subyacente siga el proceso

$$\frac{dS}{S} = rdt + \sigma(S)dW \text{ de volatilidad no constante.}$$

En estos casos se tiene que simular toda la trayectoria para calcular el payoff.

La simulación de la trayectoria debe hacerse discretizando la ecuación diferencial estocástica que gobierna la dinámica del proceso seguido por el subyacente.

Obviamente, esta discretización de la trayectoria introduce un error propio que es independiente del error de estimación de Montecarlo. Incluso, aunque se pudiera simular la trayectoria continua se incurriría en los errores propios de cualquier estimación, pero en el caso

de discretización de la ecuación diferencial estocástica se tiene además el error de discretización.

La simulación de trayectorias es muy simple como un modelo continuo con incremento fijo de tiempo. Es decir, se divide el intervalo $[0, t]$ en el que se desea la simulación en n subintervalos de consecutivos de duración Δt , $0 = t_0 \leq t_1 \leq \dots \leq t_n = t$ con $\Delta t = t_i - t_{i-1} \quad \forall i = 1, \dots, n$, siendo el mecanismo de transición $S \leftarrow S + Sr_{\Delta t} + \sigma(S)dW = Se^{r\Delta t} + \sigma(S)dW = S(1+r)^{\Delta t} + \sigma(S)dW$.

Si el proceso considerado es un proceso de Wiener, será más preciso si se utiliza la fórmula de actualización final e un intervalo, es decir, $S \leftarrow Se^{\left(r_c - \frac{1}{2}\sigma(S)^2\right)\Delta t + \sigma(S)(W(t) - W(0))}$, donde $W(t) \stackrel{d}{=} N(0, \sqrt{\Delta t})$.

Una vez calculados los valores de la trayectoria en los nodos para los valores intermedios se interpola linealmente. Así se construyen las trayectorias del subyacente en el interior del intervalo $[0, t]$.

Para cada trayectoria simulada se calcula el “pay-off” del contrato al que daría lugar. Se necesitan muchas trayectorias para poder evaluar con garantías el valor esperado del “pay-off” y su valor presente a la tasa libre de riesgo que será el valor de la opción.

Por otra parte, para hacer un intervalo de confianza o dar la precisión de la estimación, hay que tener en cuenta que $V_{call}(S) = e^{-r_c t} E_r[\text{payoff}(S)] = E_r[e^{-r_c t} \text{payoff}(S)]$. Cuando se hace la estimación mediante simulación se tiene:

$$V_{call} = e^{-r_c t} \sum_{i=1}^n \text{payoff}_i / n = \sum_{i=1}^n e^{-r_c t} \text{payoff}_i / n = \overline{e^{-r_c t} \text{payoff}}$$

y se puede hablar entonces de la precisión de esta estimación como de la varianza de la media muestral. Puesto que $V[e^{-r_c t} \text{payoff}(S)] = e^{-2r_c t} V[\text{payoff}(S)]$, se puede decir que la varianza de la estimación del valor es $S_{V_{call}}^2 = e^{-2r_c t} S_{\text{payoff}}^2 / n$, y así obtener la precisión de la estimación

$$\text{como } t_{n-1, \alpha/2} e^{-r_c t} \sqrt{\frac{S_{\text{payoff}}^2}{n}}.$$

En ocasiones puede requerir mucho esfuerzo computacional, y ser muy lento, por lo que las técnicas de reducción de la varianza para aumentar la precisión o reducir el tamaño de la simulación pueden ser muy convenientes.

La simulación es uno de los mejores procedimientos para valorar una opción que no implique ninguna decisión durante el periodo de ejercicio, como son las opciones europeas. Sin embargo, no es apropiada para valoración de opciones americanas o bermudas en las que el ejercicio se puede hacer en diferentes momentos del tiempo. Una referencia para ver cómo aplicar la simulación a este tipo de opciones es Longstaff-Schwartz(2001), sin embargo, no es lo más apropiado.

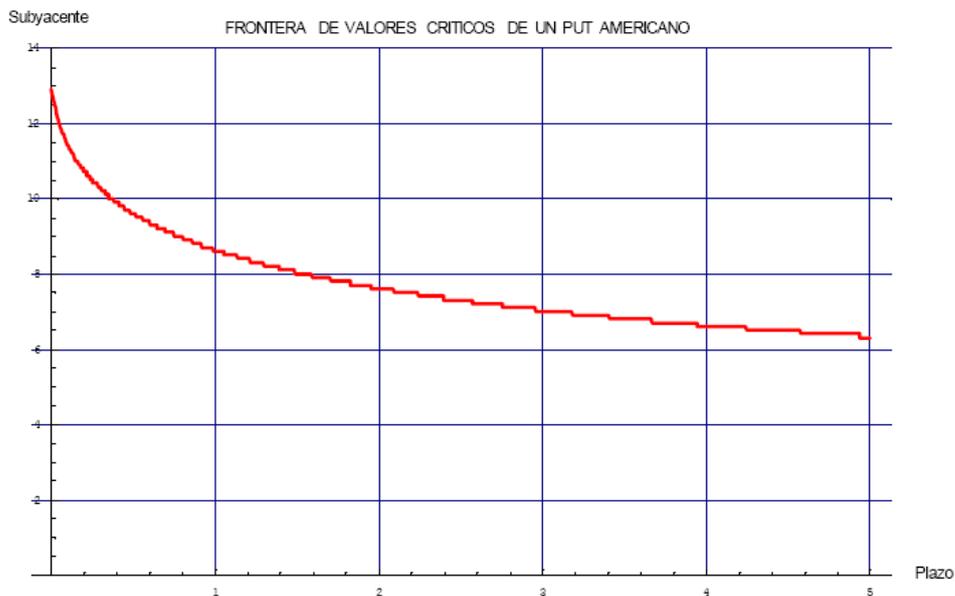
Por lo tanto, a continuación se presenta un capítulo en que se muestran métodos ya no de simulación Monte Carlo para valoración de opciones americanas.

VI.2.3 Estimación del valor de opciones con posibilidad de ejercicio anticipado mediante métodos numéricos

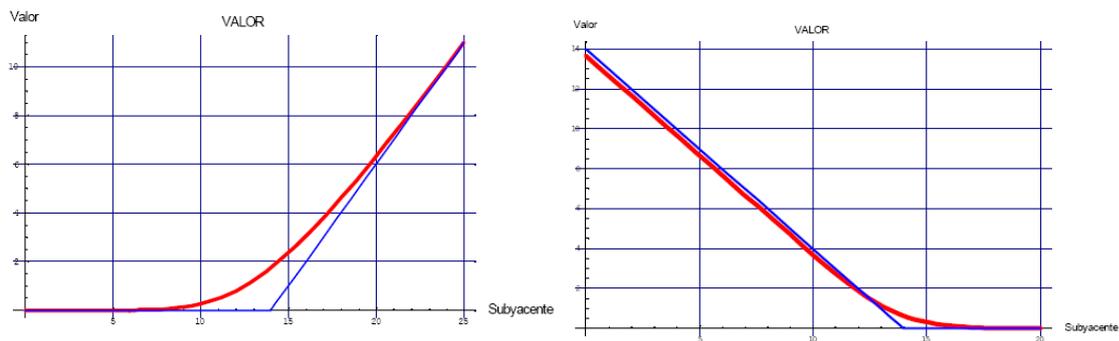
Como ya se ha dicho, la simulación no da respuesta a qué decisiones tomar sino que evalúa decisiones previamente planteadas. Por lo tanto, se presenta como una herramienta muy poderosa para valorar opciones en que no hay que tomar decisiones durante el periodo de ejecución. Sin embargo, no es una herramienta apropiada para valorar opciones que en su diseño implica tomar también alguna decisión en función del valor del subyacente durante el periodo de ejercicio, como son las opciones americanas o bermudas. En estas opciones no sólo se busca dar el valor, sino, dar además una frontera de valores críticos que determinan cuándo se debe ejercer la opción, siendo incluso este punto más relevante que la valoración de la propia opción.

Obviamente, el ejercicio óptimo varía con el subyacente y con el tiempo hasta el vencimiento, $S^*(t)$, marcando la frontera entre la conveniencia o no del ejercicio en ese momento. Haciendo variar t se obtiene la *frontera de ejercicio óptimo*, a veces también llamada *conjunto de valores críticos*.

El siguiente gráfico (de Camaño (2006)) muestra la frontera de ejercicio óptimo de una opción put americana con strike 13. Obviamente, al final del ejercicio ($t = 0$), el valor crítico es el propio strike, si el valor del subyacente es inferior al strike se ejerce la opción vendiendo a 13 (por encima de lo que vale el subyacente). Tres años antes, por ejemplo, el valor crítico es 7, con lo que interesa ejercer la opción anticipadamente si el valor del subyacente es inferior.



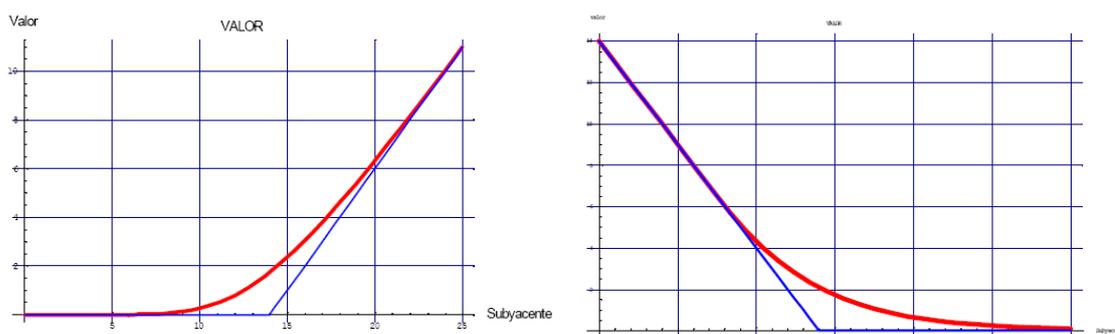
Un gráfico muy interesante para poder ver las diferencias entre el valor de opciones europeas y americanas es el gráfico de valor o del pay-off. Este gráfico representa lo que sería el pay-off en ese momento si se pudiera ejercer y el valor de la opción en función del valor del subyacente. Los dos gráficos mostrados a continuación (obtenidos de Camaño (2006)) representan los diagramas pay-off de una opción europea call y una put, respectivamente, ambas con strike 14 y a cierto plazo del vencimiento. La línea azul es el supuesto pay-off y la línea roja el valor de la opción, ambas en función del valor del subyacente. Obsérvese que la línea roja sería una sección (la que determine el tiempo hasta el vencimiento) de la superficie de valor $V(S,t)$ que se vio en la sección VI.2.1, y la línea azul la correspondiente a la fecha de vencimiento de esa misma superficie.



Obsérvese, que en la opción put el valor de la opción puede estar por debajo del pay-off que se recibiría en ese momento. Esto no puede ocurrir en una opción americana pues daría una

opción de arbitraje, bastaría con comprar la opción (pagando su valor) y ejercerla de inmediato, obteniendo un pay-off mayor sin riesgo alguno.

Los diagramas de valor de opciones americanas son los siguientes, también con strike 14 y al mismo plazo del vencimiento.



Como puede verse en la opción put, el gráfico de valor descansa sobre el pay-off, como si fuera un obstáculo, tomando un contacto suave con él. El punto de contacto es justo el valor crítico para ese plazo. En ese momento, para valores del subyacente por debajo del valor crítico se debe ejercer la opción (venta del subyacente), mientras que para valores superiores se debe mantener hasta un tiempo posterior.

Resulta obvio, en general, la siguiente relación

$$V_{\text{europea}}(S, t) \leq V_{\text{bermudas}}(S, t) \leq V_{\text{americana}}(S, t)$$

Sin embargo, se puede observar que el gráfico de la opción call es el mismo para una opción europea y americana. Sin entrar en detalles, una opción call americana nunca es óptimo ejercerla antes del vencimiento, ya que se compra el subyacente y se sigue sujeto a sus fluctuaciones (pudiendo perder más que con la opción y añadiendo los gastos financieros para poseer el activo). Por lo tanto, el valor de una opción call americana que no paga dividendos es el mismo que el de una europea. Aunque si el activo paga dividendos ya sí puede interesar ejercer anticipadamente. En efecto, ejercer la opción significa pagar el strike y cargar con el coste de financiación de anticipar el pago del strike, pero la posesión anticipada de la opción da derecho a percibir el dividendo desde ese momento. Mientras la rentabilidad por dividendo sobre el precio de la acción sea inferior, en términos absolutos, al coste de financiación del strike, no será óptimo ejercer antes del vencimiento. Este criterio no determina el valor crítico, pero marca

una condición clara, es decir, $S^*(t) \geq K \frac{e^{rt} - 1}{e^{Dt} - 1}$.

Sin embargo, en la práctica las rentabilidades por dividendos de las acciones suelen ser inferiores a las tasas de interés libres de riesgo $D < r$ y esto hace que los valores críticos se

sitúen alejados del strike, y, salvo que las opciones estén muy “in the money”, o con rentabilidades por dividendo iguales o superiores a las tasas de interés o dividendos extraordinarios, o con plazo remanente largo, se estará lejos del valor crítico y es raro que interese ejercer anticipadamente una opción call americana.

El caso de las opciones put americanas es completamente distinto al de las call americanas, tanto si pagan dividendos como si no pagan dividendos existe una frontera de valores críticos que marca el momento óptimo de ejercicio y el problema de valoración es más complejo, no siendo válida la ecuación de Black&Scholes.

En 1993 Petter Bjerksund & Gunnar Stensland publicaron un artículo en el “Journal of Business Finance and Accounting” bajo el título “American Exchange Options and a put call transformation” en el que aparece una equivalencia o relación de paridad entre los valores de una opción “call” americana y una opción “put” americana con parámetros cambiados. Esa relación es:

$$V_{call americana}(S, t, K, \sigma, r, d) = V_{put americana}(K, t, S, \sigma, d, d - r)$$

que no es una relación entre una call y su put, ya que los parámetros están cambiados, pero sirve para poder hacer una valoración en un momento dado. Es válida para americanas y europeas, siempre que sean vanilla, es decir, sin ninguna otra sofisticación.

Los métodos numéricos habituales para hacer una valoración de una opción put americana o bermudas son básicamente el método binomial y los métodos basados en diferencias finitas.

VI.2.3.1 Método binomial

El método binomial, se basa fundamentalmente en construir unas trayectorias en tiempo discreto cuya convergencia débil esté asegurada que es el movimiento browniano, y estimar el pay-off sobre estas trayectorias (en cierto modo, se podría entender como una simulación). Por lo tanto, sólo es válido asumiendo que el subyacente sigue un movimiento browniano.

El principio de invarianza de Donsker juega un papel con los procesos estocásticos similar al del teorema central del límite con las variables aleatorias. Bajo ciertas condiciones, un paseo aleatorio (“random walk”) discreto con saltos regidos por cualquier distribución que tenga varianza finita converge una vez normalizado hacia el proceso de Wiener cuando el intervalo de tiempo entre los saltos se va aproximando a cero. Los saltos pueden tener cualquier distribución pero con la condición de que tengan media nula y varianza unidad. Esta es una condición necesaria para la convergencia hacia el proceso de Wiener, si la media y la varianza no cumplen este requisito pero son finitas se producirá la convergencia igualmente pero no hacia el proceso

de Wiener sino hacia una difusión de Ito cuyos parámetros sean precisamente la media y la desviación típica de la distribución de los saltos.

La formulación del principio de invarianza sería:

“Sea una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ todas ellas independientes y con la misma distribución que la variable aleatoria X que tiene media 0 y varianza finita positiva $E[X^2] = \sigma^2 > 0$. A partir de ellas se construyen las sumas parciales $M_n = \sum_{k=1}^n X_k$. Estas sumas parciales constituyen un paseo aleatorio o “random walk” en tiempo discreto que puede transformarse en un proceso continuo mediante la interpolación lineal siguiente:

$$G_n(t) = \frac{1}{\sigma\sqrt{n}} M_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} X_{[nt]+1}$$

El principio de invarianza dice que en estas circunstancias, la sucesión de los procesos continuos $\{G_n(t)\}_{n \in \mathbb{N}}$ converge débilmente hacia el proceso de Wiener”.

Generalizando el principio de invarianza se puede comprobar que si las variables aleatorias son independientes e idénticamente distribuidas con media $\mu\delta t$ y varianza $\sigma^2\delta t$ siendo $\delta t = T/n$, los procesos continuos

$$G_n(t) = M_{[nt]} + (nt - [nt])X_{[nt]+1} \quad 0 \leq t \leq T$$

convergen débilmente en el intervalo $[0, T]$ hacia una difusión de Ito $G(t)$ cuya ecuación diferencial estocástica es $dG = \mu dt + \sigma dW$, que no es un proceso de Wiener, pero sus saltos en δt son normales de media $\mu\delta t$ y varianza $\sigma^2\delta t$.

Esta convergencia débil de los paseos aleatorios a difusiones de Ito con parámetros constantes se da cualquiera que sea la distribución de las variables aleatorias siempre que sean independientes e idénticamente distribuidas y tengan media $\mu\delta t$ y varianza $\sigma^2\delta t$.

Naturalmente, el interés no está en los paseos aleatorios aritméticos sino en los geométricos, en los que cada término se deriva del anterior multiplicando por una variable aleatoria positiva,

de modo que $S_k = S_{k-1}e^{R_k} = S_0 e^{\sum_{k=1}^n R_k}$, pero que tomando logaritmos se transforman en paseos

aleatorios aritméticos: $\ln(S_k) = \ln(S_{k-1}) + R_k = \ln(S_0) + \sum_{k=1}^n R_k$.

Los paseos aleatorios geométricos son los que convergen débilmente al GBM $dS = \mu S dt + \sigma S dW$. Dada esta convergencia débil, si se dispone de una sucesión de paseos aleatorios geométricos que convergen en distribución al GBM de riesgo neutro, puede computarse el valor esperado de cualquier variable directamente mediante la medida de probabilidad de riesgo neutro, o bien aproximando mediante la sucesión de los valores esperados de dicha variable con la sucesión de medidas de probabilidad asociadas con los paseos aleatorios de la sucesión.

La idea por lo tanto, es construirlos con unas variables aleatorias lo más sencillas posible. En concreto, se considera una variable que puede tomar dos valores posibles $u > 1$ con probabilidad p y $d = 1/u$ con $1 - p$, de modo que el subyacente en cada salto pasa a ser $S_k = S_{k-1}u$ con probabilidad p y $S_k = S_{k-1}d = S_{k-1}/u$ con probabilidad $1 - p$.

Para que se dé la convergencia requerida al GBM con riesgo neutro, el rendimiento esperado debe ser el de la tasa libre de riesgo, es decir, $e^{r\delta t} = pu + (1 - p)/u$, y la varianza $\sigma^2 \delta t$, es decir, $\sigma^2 \delta t = pu^2 + (1 - p)/u^2 - (pu + (1 - p)/u)^2$. Resolviendo este sistema se tiene:

$$u = e^{\sigma\sqrt{\delta t}} \quad p = \frac{e^{r\delta t} - 1/u}{u - 1/u}.$$

Obsérvese entonces que se obtiene un árbol binomial con los sucesivos saltos. Por otra parte, a cada nodo del árbol se llega tras una serie de subidas y bajadas, independientemente de en qué orden se hayan producido. Es decir, en la etapa n a un nodo se llega tras k subidas y $n - k$ bajadas, desde varias trayectorias, cada una con probabilidad $p^k (1 - p)^{n - k}$, y que el número de caminos que llevan a él es de $\binom{n}{k}$. Por lo tanto, la probabilidad de que sea

alcanzado ese nodo es la probabilidad de la distribución binomial $\binom{n}{k} p^k (1 - p)^{n - k}$, siendo el valor del subyacente en ese nodo $Su^k \left(\frac{1}{u}\right)^{n - k} = Su^{2k - n}$.

Con la probabilidad asociada a cada nodo se tiene definida una medida de probabilidad en el conjunto de las posibles trayectorias seguidas por el paseo aleatorio geométrico de riesgo neutro, y se puede estimar el valor del pay-off en los nodos finales de forma directa, y con ello actualizarlo con la tasa libre de riesgo y tener la valoración de una opción. Éste es el conocido método binomial.

Para una opción call europea sería

$$V_{call\ europea}(S, t) \approx e^{-rt} \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} \text{Max}(Su^{2k-n} - K, 0)$$

siendo $\delta t = t/n$ y $u = e^{\sigma\sqrt{\delta t}}$ $p = \frac{e^{r\delta t} - 1/u}{u - 1/u}$, donde estas aproximaciones convergen al valor real de la opción.

Este método, siendo válido para opciones europeas no lo es para opciones americanas, pues no siempre se ejercen al final. Una modificación inmediata para usarlo es no utilizar el valor sólo de las hojas del árbol binomial, sino, computarlo en cada nodo del árbol para poder tomar la decisión, por ejemplo, de ejecutar en ese momento si el valor esperado de etapas futuras es inferior. Este procedimiento exige ir recorriendo el árbol de las hojas a la raíz, para poder estimar el pay-off esperado, y se conoce como algoritmo binomial.

No es apropiado para opciones europeas pues aunque no maneja grandes números combinatorios, requiere mucha memoria y tiempo de computación, pero es la alternativa para opciones americanas y bermudas.

El algoritmo trata de computar el valor esperado del “pay-off” en la raíz S , pero recorriendo el árbol de las hojas hasta el vértice. Comenzando desde el nivel n -ésimo para el cual el “pay-off” se computa de acuerdo con las características del contrato y recorriendo el árbol, un paso atrás cada vez, hasta llegar al vértice. En cada nodo del nivel $t = k$ se computa el valor esperado del “pay-off” calculando la media de los valores esperados del “pay-off” en los dos nodos del nivel $k+1$ conectados con él y previamente calculados, actualizados con la tasa libre de riesgo, y se compara con la ejecución en el momento, eligiendo lo más ventajoso. El valor computado para la raíz es la estimación del valor de la opción.

VII Referencias

- Barceló, J. (1996) *Simulación de Sistemas Discretos*. Isdefe.
- Bratley, P., Fox, B.L., Schrage, L.E. (1987) *A Guide to Simulation*. Springer-Verlag.
- Camaño, A. (2006) *Opciones Financieras*
- Carter, G. and Ignall, E.J. (1975) Virtual Measures: A Variance Reduction Technique for Simulation, *Management Science*, **21**, 607-616.
- Fushimi, M. (1989) "Random Number Generation on Parallel Processors" *Proceedings of the 1989 Winter Simulation Conference*. pp 459-461.
- Geweke, J. (1989) Bayesian Inference in Econometrics using Monte Carlo Integration. *Econometrica*, **57**, 1317-1340.
- Hammersley, J.M. and Handscomb, D.C. (1964) *Monte Carlo Methods*, Methuen
- Hull, J.C. (2006) *Options, Futures and Other Derivatives*. Prentice Hall.
- Kelton, W.D., Sadowski, R.P. and Sadowski, D.A. (2002) *Simulation with Arena*. McGraw-Hill.
- Kleijnen, J. and Van Groenendaal, W. (1994) *Simulation. A Statistical Perspective*. Wiley&Sons
- Law, A.M., Kelton, W.D. (2000) *Simulation Modeling and Analysis*. McGraw-Hill.
- L'Ecuyer, P. (1999) "Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators" *Operations Research*. Vol. 47, No. 1, January-February, pp 159-164.
- L'Ecuyer, P., Simard, R., Chen, E.J., and Kelton, D. (2002) "An Object-Oriented Random-Number Package with Many Long Streams and Substreams" *Operations Research*. Vol. 50, No. 6, November-December, pp 1073-1075.
- Longstaff, F.A. and E.S. Schwartz (2001) Valuing American Options by Simulation: A Simple Least Squares Approach. eScholarship Repository of Anderson Graduate School of Management, University of California. <http://repositories.cdlib.org/anderson/fin/1-01>
- Montgomery, D.C. (1991) *Diseño y análisis de experimentos*. Ed. Iberoamericana.
- Peña, D. (1998) *Estadística. Modelos y métodos*, (vol. 2). Alianza Universidad Textos.
- Ríos-Insúa, D., Ríos-Insúa, S., Martín, J. (1997) *Simulación. Métodos y Aplicaciones*. Ra-Ma
- Rubinstein, R.Y. and Melamed, B. (1998) *Modern Simulation and Modeling*. John Wiley&Sons
- Thomson, S.K. (1992) *Sampling*. Wiley.
- Trotter, H. and Tukey, J. (1956) Conditional Monte Carlo for normal samples, in *Symposium on Monte Carlo Methods*, H. A. Meyer (ed.), Wiley, 64-79

VIII Biblioteca de problemas

VIII.1 Problemas de Métodos de Monte Carlo

PROBLEMA 1

Sea una variable aleatoria X con función de densidad y de distribución:

$$f(x) = \begin{cases} \frac{x}{8} & x \in [0,2] \\ \frac{1}{12} & x \in (2,8] \\ \frac{5}{4} - \frac{x}{8} & x \in (8,10] \\ 0 & \text{resto} \end{cases}, \quad F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{16} & x \in [0,2] \\ \frac{1}{4} + \frac{1}{12}(x-2) & x \in (2,8] \\ -\frac{x^2}{16} + \frac{5x}{4} - \frac{21}{4} & x \in (8,10] \\ 1 & \text{resto} \end{cases}$$

- Explicar cómo se generarían valores de la misma. Aplicación al caso de disponer de los números aleatorios 0.680, 0.108, 0.324, 0.814, 0.560 y 0.912.
- ¿Cómo se generarían valores de una variable aleatoria discreta que toma los valores 1, 5 y 9 con probabilidades 0.2, 0.5 y 0.3, respectivamente? Aplicarlo con los datos del apartado anterior.

PROBLEMA 2

Aplicar el método de la transformada inversa y el de aceptación y rechazo para generar variables aleatorias a partir de las siguientes funciones de densidad, y comentar qué método es el más conveniente para cada función de densidad:

$$f(x) = \begin{cases} \frac{3x^2}{2} & -1 \leq x \leq 1 \\ 0 & \text{cualquier otro valor} \end{cases}$$

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{a(1-a)} & 0 \leq x \leq a \\ \frac{1}{1-a} & a \leq x \leq 1-a \text{ para valores de } 0 < a < 1/2 \\ \frac{1-x}{a(1-a)} & 1-a \leq x \leq 1 \\ 0 & x \geq 1 \end{cases}$$

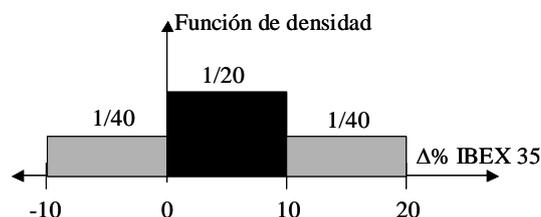
PROBLEMA 3

Un muelle circular ejerce fuerza en función del ángulo de separación θ . Dicho ángulo puede estar entre 0 y 90 grados. Su fuerza en Newton se calcula como: $K\theta^2$ expresando K en Newton/rad². El ángulo θ de separación es una variable aleatoria que tiene una función de densidad $f(\theta) = \cos(\theta)$.

- Establecer el procedimiento de la transformada inversa para muestrear la fuerza que ejerce el muelle.
- Utilizando como números aleatorios uniformes 0.32, 0.17, 0.78, 0.90 y 0.54 calcular el valor medio de la fuerza del muelle y la varianza de dicho valor medio.

PROBLEMA 4

Un fondo de inversión ofrece un interés anual equivalente al incremento porcentual en el año del índice IBEX-35. Un potencial inversor ha analizado estadísticamente el incremento porcentual de este índice y ha establecido la siguiente función de densidad:



Calcular la media muestral del incremento porcentual del IBEX-35 aplicando el método de la transformada inversa con los números pseudoaleatorios facilitados a continuación, y comparar el resultado con el valor real esperado según la distribución propuesta. Números pseudoaleatorios: 0.12, 0.28, 0.79, 0.34, 0.52, 0.37, 0.89, 0.24, 0.09, 0.93.

PROBLEMA 5

Un equipo está formado por tres componentes idénticas configuradas en paralelo y funcionando en redundancia activa, es decir, las tres al mismo tiempo y con la misma carga. El fallo de una componente es catastrófico para la misma.

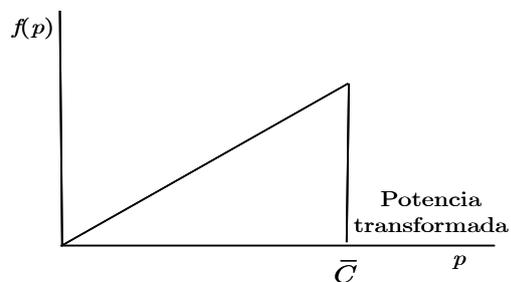
Si designamos por T a la variable duración de una componente, la probabilidad de que una cualquiera de las componentes falle antes del instante t viene dada por

$$F(t) = P(T \leq t) = 1 - e^{-\frac{t}{5}}$$

Estimar, mediante simulación y con ayuda de la secuencia de números aleatorios que se adjunta (0.51, 0.46, 0.41, 0.53, 0.21, 0.08, 0.74, 0.10, 0.65, 0.77, 0.25, 0.74, 0.89, 0.95, 0.69), la vida media del equipo.

PROBLEMA 6

Un transformador tiene una capacidad máxima de transformación \bar{C} que se distribuye según una función de densidad uniforme en el intervalo $[a, b]$. Su potencia transformada p se distribuye estadísticamente según una distribución triangular cuyo valor mínimo es 0 y cuyo valor máximo es el valor de \bar{C} y que a su vez coincide con el valor más probable de la distribución, ver figura adjunta.



- Establecer un procedimiento basado en el método de la transformada inversa para simular la potencia transformada.
- Establecer un procedimiento que utilice el método de aceptación-rechazo simple que permita simular la potencia transformada.
- Estimar por uno de los dos procedimientos la media de la potencia transformada con los valores de $a=395$ y $b=410$ a partir de la siguiente secuencia de números pseudoaleatorios: 0.435, 0.128, 0.301, 0.604, 0.321, 0.165, 0.603, 0.516, 0.321, 0.932, 0.859, 0.932.

VIII.2

PROBLEMA 7

Un producto consta de dos componentes que se producen por separado con maquinaria y personal independiente, y una vez fabricadas se ensamblan conjuntamente. En la planta de producción han estimado la distribución del tiempo que lleva fabricar cada una de las componentes, en horas, así como el tiempo de ensamblado.

El tiempo de fabricación de la primera componente se distribuye según la función de densidad

$$f(x) = \begin{cases} \frac{12-x}{32} & 4 \leq x \leq 12 \\ 0 & \text{resto} \end{cases}$$

La fabricación de la segunda componente se distribuye según la función de densidad

$$g(x) = \begin{cases} 1/6 & 6 < x < 7 \\ 1/3 & 7 < x < 9 \\ 1/6 & 9 < x < 10 \\ 0 & \text{resto} \end{cases}$$

Por último, el tiempo de ensamblado de las dos piezas sigue una distribución discreta cuya función de masa de probabilidad es: $P(1) = 0.3$ $P(1.5) = 0.5$ $P(2) = 0.2$

- 1) Dar un procedimiento para obtener un valor del tiempo total de producción del producto. Expresarlo en pseudocódigo o gráficamente.
- 2) Aplicar el procedimiento, utilizando el método de la transformada inversa para simular el tiempo de producción de cada componente, para obtener el valor medio del tiempo de producción total. Dar el valor medio y la precisión de la estimación al 95%.
- 3) Aplicar el procedimiento, utilizando el método de aceptación-rechazo para simular el tiempo de producción de cada componente, para obtener el valor medio del tiempo de producción total. Dar el valor medio y la precisión de la estimación al 95%.
- 4) Indicar cómo sería el procedimiento con variables antitéticas, y aplicarlo utilizando el método de la transformada inversa para simular el tiempo de producción de cada componente, para obtener el valor medio del tiempo de producción total. Obtener una muestra del mismo tamaño que en el apartado 2, y dar el valor medio y la precisión de la estimación al 95%.

Utilizar los siguientes valores uniformes en (0,1), secuencialmente según se necesiten: 0.01, 0.23, 0.85, 0.97, 0.34, 0.48, 0.08, 0.17, 0.71, 0.68, 0.42, 0.71, 0.16, 0.24, 0.79, 0.92, 0.84, 0.02

VIII.2 Resultados de los problemas de Métodos de Monte Carlo

RESULTADO DEL PROBLEMA 3

Datos:

- q varia entre 0 y 90 grados.
- La fórmula de la fuerza es: $F = Kq^2$ [N];
- La función de densidad de ángulo q: $f(q)=\cos(q)$; $q \in (0,90)$
- La función de distribución del ángulo q es: $F(q) = \text{sen}(q)$ $q \in (0,90)$

Generamos un número aleatorio entre 0 y 1 y luego se determina x tal que $F(x) = u$, luego $x = -\arcsen(u)$

Con los números aleatorios propuestos: 0.32, 0.17, 0.78, 0.90 y 0.54, y $x = q$:

18,66 9,79 51,26 64,15 32,68

Las diferentes fuerzas saldrán:

Fuerzas

348,2 K

95,84 K

2627 K

4117 K

1069 K

La fuerza media del muelle viene dada por:

$$\text{Media} = \frac{\sum_{i=1}^{i=5} \text{fuerzas}}{5} = K \cdot 1651,23 [\text{Newton}]$$

$$\text{La varianza del valor medio viene dada por: Valor} = \frac{s^2}{5} = \frac{\sum_{i=1}^{i=5} (xi - \bar{x})^2}{5 \times (5-1)} = K \cdot 574413;$$

VIII.2

RESULTADO DEL PROBLEMA 4

Establecemos la función de densidad y calculamos la función de distribución:

Función de densidad:

$$i(x) = \begin{cases} 1/20 \Leftrightarrow 0 \leq x \leq 10 \\ 1/40 \Leftrightarrow -10 \leq x \leq 0 \\ 1/40 \Leftrightarrow 10 \leq x \leq 20 \\ 0 \text{ resto} \end{cases}$$

La función de distribución quedará:

$$I(x) = \int_0^p i(x) dx = \begin{cases} 0 \Leftrightarrow x < -10 \\ \frac{x+10}{40} \Leftrightarrow -10 \leq x \leq 0 \\ \frac{1}{4} + \frac{x}{20} \Leftrightarrow 0 \leq x \leq 10 \\ \frac{3}{4} + \frac{x-10}{40} \Leftrightarrow 10 \leq x \leq 20 \\ 1 \Leftrightarrow x > 20 \end{cases}$$

Números pseudoaleatorios: 0.12, 0.28, 0.79, 0.34, 0.52, 0.37, 0.89, 0.24, 0.09, 0.93

Cálculo de la media muestral del incremento porcentual del IBEX-35 aplicando el método de la transformada inversa:

Regla:

Si $u \in (0, 1/4)$ entonces $u = (x+10)/40$

Si $u \in (1/4, 3/4)$ entonces $u = 1/4 + x/20$;

Si $u \in (3/4, 1)$ entonces $u = 3/4 + (x-10)/40$;

Las variables aleatorias resultantes:

U	x
0.12,	-5,2
0.28,	0,6
0.79,	11,6
0.34,	1,8

0.52, 5,4
 0.37, 2,4
 0.89, 15,6
 0.24, -0,4
 0.09, -6,4
 0.93 17,2

La media sale:
$$\text{media} = \frac{\sum_{i=1}^{i=10} xi}{10} = 4,26 \%$$

Comparando este resultado con el valor esperado según la distribución:

Gracias al gráfico adjunto se ve que sale 5 %.



La diferencia entre ambos es de 0.74 %.

RESULTADO DEL PROBLEMA 5

$$F(t) = P(T \leq t) = 1 - e^{-\frac{t}{5}} \rightarrow \text{exponencial de media 5.}$$

Estimación de la vida media del equipo:

Método de la transformada inversa: $t = -5 \ln(F)$;

Con los 15 números aleatorios sacamos 15 tiempos de duración t y tomamos estos de 3 en 3. La vida del equipo será el máximo tiempo de entre los tres; esto es porque están en paralelo, y el equipo no dejará de funcionar hasta que las tres componentes fallen.

Usando los números aleatorios:

0.51, 0.46, 0.41, 0.53, 0.21, 0.08, 0.74, 0.10, 0.65, 0.77, 0.25, 0.74, 0.89, 0.95, 0.69

fallos componentes fallos de equipo

VIII.2

1		
0.51	3.36672277	
0.46	3.88264395	4.46
0.41	4.4579906	
0.53	3.17439136	
0.21	7.80323874	12.62
0.08	12.6286432	
0.74	1.50552546	
0.1	11.5129255	11.51
0.65	2.15391458	
0.77	1.30682382	
0.25	6.93147181	6.93
0.74	1.50552546	
0.89	0.58266908	
0.95	0.25646647	1.86
0.69	1.85531841	

$$\text{Vida media} = \frac{\sum_{i=1}^{i=5} \text{tiempo}}{5} = 7,478$$

RESULTADO DEL PROBLEMA 6

Según el enunciado, vemos que la función de densidad de \bar{C} va a ser:

$$f(\bar{C}) = \frac{1}{b-a} \quad a \leq \bar{C} \leq b$$

Según la distribución triangular tenemos:

$$\int_0^{\infty} f(p) dp = 1 \rightarrow \text{Area triángulo} = h\bar{C}/2 = 1 \rightarrow h = \frac{2}{\bar{C}}$$

La función de densidad de $f(p)$ queda:

$$f(p) = \frac{2p}{\bar{C}^2} \quad 0 \leq p \leq \bar{C}$$

y la de distribución:

$$F(p) = \int_0^p f(p) dp = \begin{cases} 1 & p > \bar{C} \\ \frac{p^2}{\bar{C}^2} & 0 \leq p \leq \bar{C} \end{cases} \text{ luego } p = \bar{C}\sqrt{F(p)}$$

Método de la transformada inversa:

Tomamos dos valores aleatorios de $U[0,1]$: u_1 y u_2 ;

Con u_1 : $\bar{C} = a + (b - a)u_1$

Con u_2 : $p = \bar{C}\sqrt{u_2}$

Como tenemos 12 números aleatorios vamos a obtener 6 valores de p .

Método aceptación rechazo:

$$f(p) = \frac{2p}{\bar{C}^2} \quad 0 \leq p \leq \bar{C}$$

Vamos a necesitar tres números aleatorios: u_1 , u_2 , u_3 :

Con u_1 : $\bar{C} = a + (b - a)u_1$

Con u_2 : $p^* = \bar{C}u_2$

Con u_3 : $f(p)^* = \frac{2}{\bar{C}}u_3$

Si $f(p)^* \leq f(p^*)$ acepto p^*

Si no: no acepto.

- Resultados para $a = 395$ y $b = 410$:

Método de la transformada inversa:

Con u_1 : $\bar{C} = 395 + (410 - 395)u_1$

Con u_2 : $p = \bar{C}\sqrt{u_2}$

VIII.2

Usamos los siguientes números aleatorios:

u1,u2	\bar{C}	p
0.435	401.525	143.653951
0.128		
0.301	399.515	310.492818
0.604		
0.321	399.815	162.405621
0.165		
0.603	404.045	290.23821
0.516		
0.321	399.815	385.981989
0.932		
0.859	407.885	393.772779
0.932		

$$\text{Media} = \frac{\sum_{i=1}^{i=6} p}{6} = 281.09$$

Método de aceptación-rechazo:

u1,u2,u3	\bar{C}	p^*	$f(p)^*$	$f(p^*)$	p
0.435	401.525	51.3952	0.00063757	0.001499284	No acepto
0.128					
0.301					
0.604	404.06	129.70326	0.00158887	0.00081671	129.7
0.321					
0.165					
0.603	404.045	208.48722	0.00255417	0.001588932	No acepto
0.516					
0.321					
0.932	408.98	351.31382	0.00420069	0.00455768	No acepto
0.859					

Media: 129.7

VIII.3 Problemas de simulación de sistemas dinámicos

PROBLEMA 1

Dado un sistema de dos componentes colocadas en serie de modo que el fallo de una de ellas supone el fallo del sistema, y tal que cuando una pieza falla la otra se desconecta (no sigue funcionando hasta que la que ha fallado no sea reparada), cuyos tiempos de vida y reparación siguen las siguientes distribuciones:

Pieza 1: Tiempo Vida: Cauchy(1,0) truncada en 0

Tiempo reparación: Exponencial(2)

Pieza 2: Tiempo Vida: Uniforme(1,3)

Tiempo reparación: Exponencial(4)

- Hacer un modelo de simulación discreto con avance al próximo evento para estudiar la disponibilidad de un sistema de estas características.¹²
- Obtener la disponibilidad del sistema explicando el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, con un tiempo límite de 10 horas, y utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia para una variable)¹³:

Vida Pieza 1: 0.2, 0.9, 0.7, 0.3, 0.9, 0.8, 0.1

Reparación Pieza 1: 0.1, 0.3, 0.7, 0.6, 0.9, 0.8, 0.2

Vida Pieza 2: 0.6, 0.8, 0.2, 0.5, 0.7, 0.1, 0.9

Reparación Pieza 2: 0.4, 0.2, 0.6, 0.8, 0.6, 0.4, 0.3

PROBLEMA 2

Una empresa de envasado de productos congelados del mar dispone de dos envasadoras iguales. Los productos llegan en contenedores o lotes, según una distribución de tiempo entre las llegadas: 15 minutos con probabilidad 0.25, 30 minutos con probabilidad 0.5 y 45 minutos con probabilidad 0.25.

¹² Disponibilidad: Proporción de tiempo que el sistema está funcionando

¹³ Función densidad de la distribución Cauchy(1,0): $f(x) = \frac{1}{\pi(1+x^2)}, \forall x \in \mathbb{R}$

Cada contenedor o lote tarda en ser procesado por una envasadora un tiempo que se distribuye según una distribución exponencial de media 1 hora truncada en 45 minutos, no pudiendo ser la duración inferior a este valor.

Si un lote llega y no puede ser procesado inmediatamente, la empresa dispone de una cámara de congelación para la espera. La cámara puede alojar únicamente a un lote de productos, y el coste de funcionamiento diario (no proporcional al tiempo que se utilice) está estimado en 50 euros diarios.

Si la cámara está llena cuando llega un lote, éste habrá de ser rechazado. Se ha estimado el beneficio que produce un lote (en términos medios) en 100 euros. Se supone un día de 8 horas.

- a) Hacer un modelo de simulación discreto con avance al próximo evento para este problema, que obtenga el coste (por funcionamiento de cámaras y por pérdida de lotes) diario esperado, que admita más de una cámara de congelación.
- b) Obtener el coste de un día utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia para una variable):
Llegadas: 0.2, 0.3, 0.9, 0.7, 0.1, 0.2, 0.9, 0.2, 0.6, 0.5, 0.8, 0.3, 0.7, 0.9, 0.4
Servicios: 0.1, 0.3, 0.7, 0.6, 0.9, 0.8, 0.2, 0.5, 0.1, 0.3, 0.7, 0.2, 0.6, 0.4, 0.9

PROBLEMA 3

El taller de mantenimiento de los autobuses urbanos de Madrid consta de una zona de inspección y otra de reparación. Los autobuses, una vez acabado su turno acuden a este taller. Cada media hora un autobús finaliza su turno. En el taller, en primer lugar se inspecciona cada autobús con un tiempo distribuido según la distribución trapezoidal $f_1(x)$ que se describe más abajo. La zona de inspección es alimentada por una única cola de tipo FIFO. Históricamente, la inspección detecta que el 30 % de los autobuses requiere reparación. La zona de reparación está alimentada también por una única cola de tipo FIFO y los tiempos de reparación se distribuyen según una distribución Gamma de parámetros $p = 2$ y $a = 0.05$, truncada en 30 minutos (no puede ser inferior).

$$f_1(x) = \begin{cases} \frac{x-20}{50} & 20 < x \leq 25 \\ 0.1 & 25 < x \leq 30 \\ \frac{35-x}{50} & 30 < x \leq 35 \\ 0 & \text{resto} \end{cases}$$

- a) Obtener el número medio de autobuses que hay en el taller, explicando el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, con un tiempo límite de 4 horas, y utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable y redondear a un decimal los valores obtenidos):
- Serie 1: 0.5, 0.9, 0.3, 0.7, 0.9, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2
- Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.2, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6
- Serie 3: 0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5
- b) Hacer un modelo de simulación discreto con avance al próximo evento para estudiar el número medio de autobuses que hay en el taller de mantenimiento

PROBLEMA 4

Una empresa fabrica dos tipos de productos que han de ser procesados por un único empleado. El tiempo de fabricación del primer tipo de producto sigue una distribución normal de media 4 horas y desviación típica 1, y el del segundo tipo de producto sigue una distribución triangular simétrica entre 2 y 4 horas, es decir, con función de densidad:

$$f(x) = \begin{cases} x - 2 & x \in (2, 3) \\ 4 - x & x \in (3, 4) \end{cases}$$

Los pedidos son unitarios, el tiempo que transcurre entre pedidos es uniforme entre 3 y 5 horas, y el 35 % de ellos son pedidos de tipo 1.

- a) Obtener el número medio de pedidos sin atender al finalizar el día, explicando el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, con un tiempo límite de 40 horas, y utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable y redondear a un decimal los valores obtenidos):
- Serie 1: 0.5, 0.9, 0.3, 0.7, 0.9, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2
- Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.2, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6
- Serie 3: 0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5
- Serie 4: 0.2, 0.7, 0.5, 0.3, 0.1, 0.9, 0.5, 0.6, 0.2, 0.8, 0.7, 0.4, 0.5, 0.6, 0.9, 0.3, 0.1, 0.4
- b) Hacer un modelo de simulación discreto con avance al próximo evento para estudiar el número medio de pedidos pendientes de ser servidos en la fábrica.

PROBLEMA 5

Se considera un taller de reparaciones con dos mecánicos. El primer mecánico cobra a razón de 30 €/h y puede reparar a razón de 5 máquinas por hora. El segundo mecánico cobra 50 €/h

VIII BIBLIOTECA DE PROBLEMAS

pero puede reparar a razón de 8 máquinas por hora. Ambos sueldos se pagan con independencia de que los trabajadores estén ocupados u ociosos. Los clientes estiman que una máquina averiada les representa un coste de 80 €/h y el taller lo asume como coste por unidad de tiempo de espera por cliente. Se supone que las máquinas llegan al taller según un proceso de Poisson de parámetro 4 máquinas por hora y que el tiempo de reparación sigue una distribución exponencial.

- a) Escribir un modelo que permita evaluar el coste de cada mecánico (un único programa que evalúe el coste de cada uno por separado).
- b) A la vista de los siguientes resultados ¿Qué trabajador resulta más económico para la empresa? Razonar la respuesta.

Relative clock	720.00	Absolute clock	720.00		
	(1)	(2)	(3)	(4)	(5)
Facility	Average	Number	Average		
	utilization	entries	time/trans		
MECAN1	.80	2869	.20		
MECAN2	.51	2892	.13		
	(1)	(2)	(3)	(4)	(5)
Queue	Maximum	Average	Total	Zero	Percent
(AD set)	contents	contents	entries	entries	zeros
TLLER1	19	3.73	2869	0	.00
MECAN1	18	2.93	2869	566	19.73
TLLER2	11	.98	2892	1	.03
MECAN2	10	.47	2892	1455	50.31
	(6)	(7)	(8)		
Queue	Average	\$Average	Current		
(AD set)	time/trans	time/trans	contents		
TLLER1	.94	.94	1		
MECAN1	.74	.92	0		
TLLER2	.24	.24	0		
MECAN2	.12	.23	0		

\$Average time/trans=average time/trans excluding zero entries

PROBLEMA 6

A una centralita llaman clientes de dos tipos, con tiempos entre llamadas distribuidos según una distribución Exponencial de parámetro 1 (minutos) para el primer tipo de clientes y uniforme entre 2 minutos y 5 minutos para el segundo tipo. La centralita tiene varias líneas con unos amables caballeros que atienden cada una de ellas. Si al recibir una llamada todas las líneas están ocupadas, salta una espantosa musiquilla con la que se suponen entretienen al

cliente de modo que éste queda pacientemente esperando a que su llamada sea atendida, lo cual se hace por riguroso orden de llamada.

Una vez que la llamada del cliente es atendida, se le pasa la llamada a unas agradables señoritas que amablemente resuelven la razón de la llamada, escuchando una segunda horrorosa música si todas las señoritas están atendiendo otras llamadas hasta que le toque el turno de ser atendido por éstas.

La distribución del tiempo que cada caballero tarda en atender la llamada se distribuye según una uniforme entre 3 y 5 minutos, y la distribución del tiempo que las señoritas tardan en resolver el problema uniforme entre 5 y 15 minutos.

Escribir un modelo/programa para estudiar en una jornada de 8 horas el número de líneas que debe tener la centralita y de señoritas que resuelven el problema, dando para el diseño óptimo el tiempo medio que pasa un cliente escuchando la espantosa primera musiquilla, el tiempo medio que pasa un cliente escuchando la horrorosa segunda música, y una tabla de frecuencias y la media del tiempo que pasa cada tipo de clientes “colgado del teléfono”, además del grado de ocupación de cada uno de los amables caballeros y de las eficientes señoritas.

PROBLEMA 7

En una fábrica se producen dos tipos de producto según peticiones. La demanda de los productos es unitaria siguiendo el tiempo entre pedidos una distribución exponencial de media 2 horas para el primer producto, y de media 30 minutos para el segundo producto.

El proceso de producción es diferente para cada tipo de producto, involucrando a 5 tipos de máquinas y con diferente tiempo de proceso en cada una de ellas según el producto de que se trate (todo en horas):

	M1	M2	M3	M4	M5
P1	U(0.5,1)		U(0.75,1.25)	U(0.3,0.5)	U(0.5,0.8)
P2	U(0.5,1)	U(1,1.5)	U(0.75,1.25)		U(0.75,1.25)

Diseñar el número de máquinas que se debe poner en la planta suponiendo producción continua, y para esta configuración dar el tiempo medio desde que llega la petición de un producto hasta que éste es terminado (para cada tipo de producto) con el fin de poder dar una aproximación del tiempo en que podrá ser servido, el máximo número de cola que se forma delante de cada tipo de máquinas para saber cuanto espacio es necesario reservar, así como el tamaño medio de las colas.

PROBLEMA 8

Una compañía ferroviaria pinta sus propios vagones, según se vayan necesitando, en sus propios talleres, donde se pinta a mano de uno en uno con una velocidad que se distribuye según una exponencial de media uno cada 4 horas y un coste anual de 4 millones de u.m. Se ha determinado que los vagones pueden llegar según un proceso de Poisson de media uno cada 5 horas. Además el coste por cada vagón que no está activo es 500 u.m/hora.

Se plantean otras dos posibilidades. Una es encargar dicho trabajo a una empresa de pintura que lo haría con aerosol con el consiguiente ahorro de tiempo. Sin embargo el presupuesto para esta segunda alternativa es de 10 millones de u.m. anuales. En este caso el proceso se aproxima a uno de Poisson con una tasa de uno cada 3 horas. La otra opción es poner otro taller exactamente igual al que hay actualmente, con igual tasa de servicio y coste anual, que permita pintar dos vagones a la vez. En todos los casos el trabajo se considera ininterrumpido, esto es, se trabajan $24 \times 365 = 8760$ horas anuales.

A la vista de los resultados de la simulación:

- a) ¿Cuál de los tres procedimientos es preferible?
- b) Si se deseara conocer la distribución del tiempo que pasan actualmente los vagones inactivos, ¿qué información sería relevante? ¿Cómo la utilizarías? Con la información actual ¿qué distribución de probabilidad conjeturarías?
- c) ¿Cuál es el nivel de utilización de los recursos en cada caso?

```
SIMULATEtall3      GENERATE 5*FN$XPDIS  GENERATE 5*FN$XPDIS  GENERATE 8760
STORAGE           2QTABLE ARRIVE OPT2      ARRIVE OPT2          TERMINATE 1
OPT1,0,5,15      SEIZE aer,Q        SEIZE aer,Q          START 1
GENERATE          ADVANCE 3*FN$XPDIS  ADVANCE 3*FN$XPDIS  END
5*FN$XPDISARRIVE  RELEASE aer        RELEASE aer
OPT1SEIZE        DEPART OPT2       DEPART OPT2
tall1,QADVANCE   TERMINATE         TERMINATE
4*FN$XPDIS
RELEASE tall1
DEPART OPT1
TERMINATE
```

Facility	(1) Average utilization	(2) Number entries	(3) Average time/trans
TALL1	.77	1682	4.02
AER	.62	1754	3.08

Storage contents	Capacity utilization	Average	Average time/trans	Entries	Average
.37	1711	3.76	TALL3	2	.73

Storage contents	(6) Current contents	(7) Maximum contents

TALL3	0	2				
Average (AD set)	(1) Total contents	(2) Zero contents	(3) Percent entries	(4) entries	(5) Queue zeros	Maximum
OPT1	19	3.04	1682	0	.00	
TALL1	18	2.26	1682	418	24.85	
OPT2	13	1.53	1754	0	.00	
AER	12	.92	1754	688	39.22	
OPT3	7	.84	1711	0	.00	
TALL3	5	.10	1711	1355	79.19	

Queue (AD set)	(6) Average time/trans	(7) \$Average time/trans	(8) Current contents
OPT1	15.81	15.81	0
TALL1	11.79	15.69	0
OPT2	7.65	7.65	1
AER	4.57	7.52	0
OPT3	4.28	4.28	0
TALL3	.52	2.50	0

\$Average time/trans=average time/trans excluding zero entries

Table

Entries in table	(1)	(2)	(3)	(4)	
1682	Mean argument	Standard deviation	Sum of arguments		
Range	15.81	14.87	26592.06	Cumulative	Cumulative
frequency	Observed percentage	Per cent remainder	-	0	0
.00	100.00				
.01 -	5	436	25.92	25.92	74.08
5.01 -	10	358	21.28	47.21	52.79
10.01 -	15	232	13.79	61.00	39.00
15.01 -	20	161	9.57	70.57	29.43
20.01 -	25	121	7.19	77.76	22.24
25.01 -	30	103	6.12	83.89	16.11
30.01 -	35	88	5.23	89.12	10.88
35.01 -	40	56	3.33	92.45	7.55
40.01 -	45	35	2.08	94.53	5.47
45.01 -	50	30	1.78	96.31	3.69
50.01 -	55	19	1.13	97.44	2.56
55.01 -	60	11	.65	98.10	1.90
60.01 -	65	10	.59	98.69	1.31
Overflow	22	1.31	100.00		.00
(5) Average value of overflow		72.84			

PROBLEMA 9

El departamento de contabilidad de una empresa tiene un sistema informático con varios ordenadores en un mismo dominio, un ordenador que actúa como servidor de impresión (que procesa los trabajos por riguroso orden de llegada sin que sean establecidas prioridades en los trabajos) y una impresora (con un buffer en principio “ilimitado”). Los miembros del departamento han elevado una solicitud (queja) a la dirección para que sea mejorado este servicio de impresión, aduciendo que la máquina al final de mes (instante de cierre contable) está sobresaturada. La dirección desea hacer un modelo de simulación para analizar el uso de la impresora, es decir, para obtener el factor de utilización de la máquina.

Tras realizar las observaciones oportunas en esta época de final de mes, y analizar los datos obtenidos llega a las siguientes conclusiones:

- El tiempo que pasa entre órdenes de impresión que llegan al servidor se distribuye según una exponencial de media 5 minutos.
- El tiempo que el servidor tarda en procesar la tarea solicitada sigue una distribución trapezoidal cuya *función de densidad* es (minutos)

$$f(x) = \begin{cases} \frac{4}{3}x & 0 < x < 0.5 \\ 2/3 & 0.5 < x < 1.25 \\ \frac{2}{3}(2.25 - x) & 1.25 < x < 2.25 \\ 0 & \text{resto} \end{cases}$$

- El tiempo que tarda en imprimir los trabajos la impresora, se distribuye según la *función de distribución* (minutos)

$$G(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{12} & 1 < x < 10 \\ \frac{x+5}{20} & 10 < x < 15 \\ 1 & x > 15 \end{cases}$$

- a) Obtener el factor de utilización de la impresora a final de mes simulando media hora, mediante una traza, explicando qué son todos los elementos del modelo (variables de estado, datos, hipótesis, eventos, ...) y el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable y *redondear a un decimal* los valores obtenidos):
- Serie 1: 0.5, 0.9, 0.3, 0.7, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2, 0.9
- Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.3, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6
- Serie 3: 0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5
- b) Hacer un modelo para este problema (con distribuciones uniformes de *tiempos de servicio* con límites los mismos de las distribuciones del enunciado), que sea lo mínimo para ver el factor de utilización de la impresora.

PROBLEMA 10

Se considera un taller de reparaciones con un robot. Se considera la posibilidad de sustituir éste por otro robot. Ambos robots son diferentes, con diferentes costes. En la siguiente tabla se recogen los datos de cada uno de los robots, entendiéndose por costes operativos los costes por unidad de tiempo que el robot está realizando alguna reparación y por costes de amortización costes por unidad de tiempo, esté reparando o no el robot.

	Costes amortización (€/h)	Costes operativos (€/h)
Robot 1	1000	2000
Robot 2	1500	4000

El primer robot puede reparar a razón de 6 máquinas por hora y el segundo puede reparar a razón de 7 máquinas por hora.

Cuando una máquina se encuentra en el taller de reparación se considera un coste por unidad de tiempo por pérdida de productividad estimado en 8000 €/h y el taller lo asume como coste por unidad de tiempo y por cliente. Se supone que las máquinas llegan al taller según un proceso de Poisson de parámetro 5 máquinas por hora y que el tiempo de reparación con que reparan los robots sigue una distribución exponencial.

Además los robots se averían, siendo

$$f(x) = \begin{cases} \frac{32}{3}(x - 0.75) & 0.75 \leq x \leq 1 \\ \frac{16}{3}(1.5 - x) & 1 \leq x \leq 1.5 \\ 0 & \text{resto} \end{cases}$$

la distribución del tiempo en horas hasta que se averían y 15 minutos el tiempo que tarda en ser reparado un robot. Cuando una máquina se estropea se interrumpe lo que esté haciendo, si es que está activa en ese momento, continuando el proceso una vez reparado en el punto en que se dejó el proceso.

- a) Obtener el coste por hora del primer robot, mediante una traza que simule hasta 2 horas (no incluir eventos posteriores a 2 horas), explicando cuáles son todos los elementos del modelo (variables de estado, datos, hipótesis, eventos, ...) y el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, utilizando los siguientes valores

obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable en el orden en que se piden en la hoja de soluciones y *redondear a 2 decimales* los valores obtenidos):

Serie 1: 0.5, 0.9, 0.3, 0.7, 0.2, 0.1, 0.6, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2, 0.9, 0.8

Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.1, 0.3, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6, 0.7

Serie 3: 0.9, 0.1, 0.8, 0.6, 0.2, 0.5, 0.7, 0.4, 0.1, 0.6, 0.3, 0.7, 0.3, 0.1, 0.1, 0.4, 0.3, 0.1

- b) Hacer un modelo para el problema del apartado anterior, considerando distribución uniforme del tiempo de vida con límites los mismos de la distribución del enunciado, y considerando que la ruptura de la máquina no es tal, sino un proceso de mantenimiento que cuando se lanza espera para ser procesado a que todos los trabajos previos ejecutándose o en espera hayan acabado.

PROBLEMA 11

Juan Pérez está estudiando la posibilidad de abrir un taller de lavado de vehículos. Los vehículos que solicitan el servicio pueden ser dos tipos: turismos y camiones. El tiempo que se tarda en lavar un turismo es de 4 minutos, y el tiempo de un camión es el doble que el requerido por los turismos, siendo el porcentaje de turismos del 65 %. Supóngase que no hubiera limitación de espacio y que utiliza una máquina rápida. Para asegurar que el lavado es correcto un 25 % de los vehículos (elegidos de forma aleatoria, no con un orden) que terminan un servicio han de pasar de nuevo por el servicio de lavado para recibir otro lavado cuya duración

sigue una distribución con función de densidad $f(x) = \begin{cases} \frac{12-x}{32} & 4 \leq x \leq 12 \\ 0 & \text{resto} \end{cases}$. Estos

vehículos tienen prioridad no interruptora respecto a los del lavado normal. Por otra parte, el tiempo entre dos llegadas de vehículos sigue una distribución Exponencial de media 7 minutos.

- a) Hacer un modelo de simulación de eventos discretos con incremento de tiempo al próximo evento para obtener los ingresos esperados durante un tiempo máximo T si los turismos pagan 3 euros y los camiones 5, definiendo todos los elementos que intervienen.
- b) Obtener los ingresos durante 45 minutos del funcionamiento del taller mediante una traza, explicando el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable en el orden en que se piden en la hoja de soluciones y *redondear a 2 decimales* los valores obtenidos):

Serie 1: 0.5, 0.9, 0.3, 0.6, 0.2, 0.3, 0.8, 0.1, 0.9, 0.7, 0.6, 0.1, 0.5, 0.1, 0.3, 0.5, 0.1, 0.8

Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.1, 0.3, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6, 0.7

Serie 3: 0.9, 0.2, 0.8, 0.6, 0.2, 0.5, 0.7, 0.4, 0.1, 0.6, 0.3, 0.7, 0.3, 0.5, 0.1, 0.4, 0.3

PROBLEMA 12

Sea una planta de producción con un puesto al que llegan dos tipos de producto, P1 y P2. El tiempo entre llegadas de los productos se considera que sigue una distribución F1, de las cuales el 65 % son productos de tipo P1. Los productos llegan en un orden aleatorio. En el puesto hay un robot que tarda en procesar los productos de tipo P1 un tiempo distribuido según una distribución G1 y los productos de tipo P2 según una distribución G2. El robot tiene dos modos de configuración, uno para cada tipo de producto, de forma que para cambiar de modo de configuración (pasar de procesar un producto de un tipo a procesar uno de otro tipo), el robot requiere un tiempo de ajuste de Taj minutos. Por esta razón, el orden en que se procesan los productos es tal que si el robot está procesando un producto de un tipo a continuación procesará otro del mismo tipo, hasta que no le queden más, es decir, hasta que al acabar un producto no queden más de ese tipo en la cola. Inicialmente, el modo en que está configurado el robot es el modo 1, es decir, para procesar productos de tipo P1.

- Hacer un modelo de simulación de eventos discretos con incremento de tiempo al próximo evento para obtener el número de productos procesados de cada tipo durante un tiempo T y el número de productos de cada tipo que quedan sin procesar al final, definiendo todos los elementos que intervienen.
- Obtener durante 100 minutos los productos procesados de cada tipo, y el número de productos que quedan sin procesar de cada tipo al final, siendo:

$$f(x) = \begin{cases} 0.1345e^{-\frac{x-10}{16}} & 10 < x < 20 \\ 0 & \text{resto} \end{cases}$$

función de densidad de la distribución F1, G1 determinista de valor 5 minutos, G2 determinista de valor 30 minutos, y el tiempo de ajuste Taj de 5 minutos. Explicar el procedimiento utilizado para generar las variables aleatorias que aparecen en el modelo, utilizando los siguientes valores obtenidos independientemente y con distribución uniforme en el intervalo (0,1) (cada secuencia una variable en el orden en que se piden en la hoja de soluciones y redondear a 2 decimales los valores obtenidos):

Serie 1: 0.5, 0.9, 0.3, 0.6, 0.2, 0.3, 0.8, 0.1, 0.9, 0.7, 0.6, 0.1, 0.5, 0.1, 0.3, 0.5, 0.1, 0.8

Serie 2: 0.8, 0.1, 0.9, 0.6, 0.2, 0.3, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6, 0.7, 0.3

PROBLEMA 13

El servicio de entrega de las declaraciones de la renta en la Agencia Tributaria consiste en dos ventanillas, la primera para comprobar que la declaración es correcta y la segunda para pagar o recibir el importe de la declaración. Una vez que la declaración es verificada en la ventanilla primera toda persona se pone en cola en la segunda. Una vez saldada su cuenta con Hacienda la persona se va de la Agencia Tributaria. Los declarantes llegan a la Agencia según una distribución exponencial de tiempo entre llegadas con media de 1 minuto. Los tiempos de servicio de las respectivas ventanillas siguen una distribución discreta. El tiempo de servicio de la primera es de 0.6 minutos con probabilidad 0.2, 0.7 minutos con probabilidad 0.5 y 0.8 minutos con probabilidad 0.3. El tiempo de servicio de la segunda es de 0.8 minutos con probabilidad 0.1, 0.9 minutos con probabilidad 0.7 y 1 minuto con probabilidad 0.2.

- a) Hacer un modelo de simulación que simule el proceso estocástico que ocurre en la Agencia Tributaria durante 10 minutos, con el objetivo de estimar el tiempo medio pasado por los clientes en dicha agencia y el número medio de personas en cada cola y en el sistema. Suponer que las ventanillas están vacías al comienzo.
- b) Representar en una tabla el avance del tiempo de simulación, los eventos que se producen y su tipo, y el valor instantáneo del número de personas en cada cola de cada ventanilla y siendo atendidas.

Para representar el comportamiento de las variables aleatorias se han extraído estas tres series independientes y uniformes de números pseudoaleatorios distribuidos entre 0 y 1 suficientemente largas que se utilizarán para determinar cada una de las variables aleatorias consecutivamente.

Serie 1: 0,76 0,06 0,16 0,79 0,54 0,59 0,17 0,88 0,5 0,93 0,33 0,64 0,51 0,73 0,23 0,62
Serie 2: 0,12 0,03 0,3 0,38 0,79 0,46 0,19 0,5 0,14 0,6 0,44 0,79 0,01 0,99 0,63 0,37
Serie 3: 0,74 0,73 0,42 0,91 0,02 0,49 0,71 0,62 0,98 0,59 0,05 0,17 0,24 0,78 0,84 0,4

VIII.4 Resultados de los problemas de simulación de sistemas dinámicos

RESULTADO DEL PROBLEMA 2

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

➤ Variables de estado:

El objetivo de la simulación va a ser estudiar el coste diario esperado en una empresa de envasado; se va a tener en cuenta que el número de cámaras de congelación puede ser más de una (para ello, utilizamos el parámetro K). Para saber el coste diario, al final de cada día deberemos saber el número de lotes rechazados. Para ello vamos a usar las siguientes variables de estado:

$N1(t)$ = Número de lotes que hay en la envasadora 1ª .

$N2(t)$ = Número de lotes que hay en la envasadora 2ª .

$NR(t)$ = Número de lotes que se rechazan .

$NE(t)$ = Número de envases que se congelan en las K camaras disponibles.

Nota: El número total de clientes en nuestro sistema va a ser $N1+N2+NE$.

➤ Eventos:

Los eventos que se pueden dar en nuestro sistema tal que produzcan un cambio en el mismo son:

- Llegada de un lote.
- Servicio de un lote en envasadora 1ª.
- Servicio de un lote en envasadora 2ª.

➤ Mecanismo de transición:

Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:

$$\begin{array}{l}
 - \text{ Llegada de un lote.} \\
 \left\{ \begin{array}{l}
 N1 \rightarrow N1 + 1 \Leftrightarrow N1 = 0 \\
 N2 \rightarrow N2 + 1 \Leftrightarrow N2 = 0 \cdot N1 = 1 \\
 NE \rightarrow NE + 1 \Leftrightarrow N2 = 1 \cdot N1 = 1 \cdot NE = 0 \\
 NR \rightarrow NR + 1 \Leftrightarrow N2 = 1 \cdot N1 = 1 \cdot NE = 1
 \end{array} \right.
 \end{array}$$

- Servicio de un lote en envasadora 1ª. $N1 \rightarrow N1-1$ o $NE \rightarrow NE-1$
- Servicio de un lote en envasadora 2ª. $N2 \rightarrow N2-1$ o $NE \rightarrow NE-1$

➤ Datos:

- Distribución de tiempos entre llegadas al sistema (Discreta).
- Distribución de los tiempos de servicio en ambas envasadoras (Exponencial)
- TMAX = tiempo máximo de simulación. (8 horas = 480 minutos).

➤ Otras variables:

- **TM:** Reloj de simulación.
- **DL:** Tiempo entre llegadas = Discreta.
- **DS1:** Tiempo de servicio envasadora1 = Exponencial.
- **DS2:** Tiempo de servicio envasadora2 = Exponencial.
- **TL:** Instante de la próxima llegada.
- **TS1:** Instante del próximo final de servicio en envasadora1
- **TS2:** Instante del próximo final de servicio en envasadora2

MODELO:

Programa principal:

0.- Inicialización

$N1 = 0; N2 = 0; NR = 0; NE = 0; TM = 0; TS1 = \text{inf.}; TS2 = \text{inf.};$
 Generar DL; TL = DL

1.- Actualización del reloj de simulación:

$TM = \min (TL, TS1, TS2);$
 $N = N1 + N2 + NE;$

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL$. Llamar a subrutina: *Llegada a sistema.*
 Si $TM = TS1$. Llamar a subrutina: *Servicio1.*
 Si $TM = TS2$. Llamar a subrutina: *Servicio2.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar:
 $N1, N2, NR. Coste = 5000 \times k + NR \times 10000$

Subrutinas

Llegada a sistema

- 1.- Si $N1 = 0$, $N1=N1+1$. Generar $DS1$, poner $TS1=TM+DS1$
Si no
Si $N2= 0$, $N2= N2+1$. Generar $DS2$ poner $TS2=TM+DS2$
Si no,
Si $NE < k$ Entonces $NE=NE+1$
Si no, $NR= NR + 1$;
- 2.- Generar DL , poner $TL=TM+DL$.
- 3.- Volver a Programa principal.

Servicio1

- 1.- Si $NE>0$ entonces $NE=NE-1$; Generar $DS1$, poner $TS1 =TM +DS1$,
Si no, $N1=N1-1$, $TS1=inf$.
- 3.- Volver a Programa principal.

Servicio2

- 1.- Si $NE>0$ entonces $NE=NE-1$; Generar $DS2$, poner $TS2 =TM +DS2$,
Si no, $N2=N2-1$, $TS2=inf$.
- 3.- Volver a Programa principal.

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Los tiempos de llegada al sistema: distribución discreta.
Probabilidad de 0.25 que sean 15 minutos: Si $u<0.25$, $x \leftarrow 15$
Probabilidad de 0.5 que sean 30 minutos: Si $0.25<u<0.75$, $x \leftarrow 30$
Probabilidad de 0.25 que sean 45 minutos: Si $u>0.75$, $x \leftarrow 45$
- Valores para DL con la serie 1:
0.2, 0.3, 0.9, 0.7, 0.1, 0.2, 0.9, 0.2, 0.6, 0.5, 0.8, 0.3, 0.7, 0.9, 0.4, 0.8
- Llegadas al servicio1 y 2: distribución exponencial de media 1 minuto, truncada en 45 minutos: $x=-Ln(u)$;
- Valores para $DS1$ y $DS2$ con la serie 2:
0.1, 0.3, 0.7, 0.6, 0.9, 0.8, 0.2, 0.5, 0.1, 0.3, 0.7, 0.2, 0.6, 0.4, 0.9 0.47 0,19 0,074

TRAZA: ($k=1$)

DL: Tiempo de llegada a la ventanilla 1:

15 30 45 30 15 15 45 15 30 30 45 30 30 45 30 45

DS1: Tiempos de servicio1 y 2:

138 72 nulo nulo nulo nulo 96 nulo 138 72 nulo 96 nulo 54 nulo 45 99 156

Nota: Usaremos los números aleatorios para ambos servicios.

Nº even.	TM	Evento	N1	N2	NR	NE	TL	TS1	TS2
0	0	Inicio	0	0	0	0	15	∞	∞
1	15	Llegada	1	0	0	0	45	153	∞
2	45	Llegada	1	1	0	0	90	153	117
3	90	Llegada	1	1	0	1	120	153	117
4	117	Servicio2	1	1	0	0	120	153	213
5	120	Llegada	1	1	0	1	135	153	213
6	135	Llegada	1	1	1	1	150	153	213
7	150	Llegada	1	1	2	1	195	153	213
8	153	Servicio1	1	1	2	0	195	291	213
9	195	Llegada	1	1	2	1	210	291	213
10	210	Llegada	1	1	3	1	240	291	213
11	213	Servicio2	1	1	3	0	240	291	285
12	240	Llegada	1	1	3	1	270	291	285
13	270	Llegada	1	1	4	1	315	291	285
14	285	Servicio2	1	1	4	0	315	291	381
15	291	Servicio1	0	1	4	0	315	∞	381
16	315	Llegada	1	1	4	0	345	345	381
17	345	Llegada	1	1	4	1	375	345	381
18	345	Servicio1	1	1	4	0	375	390	381
19	375	Llegada	1	1	4	1	420	390	381
20	381	Servicio2	1	1	4	0	420	390	480
21	390	Servicio1	0	1	4	0	420	546	480
22	420	Llegada	1	1	4	0	450	546	480
23	450	Llegada	1	1	4	1	495	546	480
24	480	Servicio2	1	1	4	0			

La solución final y objetivo del problema, será:

Al cabo de 8 horas se tiene que

$$\text{Coste} = 5000 \times k + NR \times 10000 = 5000 \times 1 + 4 \times 10000 = 45000 \text{pts} = 272\text{€}.$$

RESULTADO DEL PROBLEMA 3

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

➤ Variables de estado:

El objetivo de la simulación va a ser estudiar el número de autobuses que hay en el sistema, formado por las zonas de inspección y de reparación (vamos a saber cuál es el nº de autobuses en cada zona en cada momento):

$N1(t)$ = Número de clientes que hay en el servidor 1 (zona de inspección).

$N2(t)$ = Número de clientes que hay en el servidor 2 (zona de reparación).

Nota: El número total de clientes en nuestro sistema va a ser $N1 + N2$.

➤ Eventos:

Los eventos que se pueden dar en nuestro sistema tal que produzcan un cambio en el mismo son:

- Llegada de un autobús al servidor 1.
- Servicio de un autobús en el servidor 1, que implica la llegada de autobús al servidor 2 o la salida definitiva del sistema después de servicio 1
- Servicio de un autobús en el servidor 2 (lo que equivale a irse definitivamente del sistema).

➤ Mecanismo de transición:

Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:

- Llegada de un autobús al servidor 1 : $N1 \rightarrow N1+1$.
- Servicio de un autobús del servidor 1. $N1 \rightarrow N1-1$. y con probabilidad p es la llegada de un autobús al servidor 2 $N2 \rightarrow N2+1$, o la salida definitiva del sistema después de servicio 1.
- Servicio de un autobús en el servidor2: $N2 \rightarrow N2-1$.

- Datos:
 - Distribución de tiempos entre llegadas a la zona de inspección (la llegada se produce cada 30 minutos).
 - Distribución de los tiempos de servicio en la zona de inspección 1 (distribución trapezoidal).
 - Llegada de autobuses a la zona de reparación (discreta), p .
 - Distribución de los tiempos de servicio en la zona de reparación (distribución Gamma de parámetros $p = 0,2$ y $a = 0,05$, truncada en 30 minutos).
 - TMAX = tiempo máximo de simulación. (4 horas =240 minutos).
- Otras variables:
 - **TM:** Reloj de simulación.
 - **DL1:** Tiempo entre llegadas al servidor 1 = Cte.
 - **DS1:** Tiempo de servicio ventanilla1 = Trapezoidal.
 - **DS2:** Tiempo de servicio ventanilla2 = Gamma.
 - **DP:** Tipo de servicio después de salir de servidor1 (llegada a servidor 2 o salida definitiva).
 - **TL1:** Instante de la próxima llegada.
 - **TS1:** Instante del próximo final de servicio en ventanilla1.
 - **TS2:** Instante del próximo final de servicio en ventanilla2.
 - **SUMA:** contador acumulado suma de áreas de autobuses en el sistema por tiempo de permanencia.
 - **TANT:** variable auxiliar.
 - **NANT:** variable auxiliar.

MODELO:

Programa principal:

0.- Inicialización

$N1 = 0; N2 = 0; TM = 0; TS1 = \text{inf.}; TS2 = \text{inf.};$

$DL1 = 30; TL = DL1; SUMA = 0;$

1.- Actualización del reloj de simulación: $TM = \min (TL, TS1, TS2);$

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL1$. Llamar a subrutina: *Llegada a servidor 1.*

Si $TM = TS1$. Llamar a subrutina: *Servicio1.*

Si $TM = TS2$. Llamar a subrutina: *Servicio2.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar: $N1, N2$.

$N = N1 + N2$;

$SUMA/TM$

Subrutinas

Llegada a servidor1

1.- $N1=N1+1$:

2.- Generar DL , poner $TL1= TM+DL$.

Si $N1=1$, Generar $DS1$, poner $TS1=TM+DS1$

3.- $SUMA = SUMA + NANT \times (TM-TANT)$

4.- $NANT=N1+N2$, $TANT=TM$

5.- Volver a Programa principal.

Servicio1

1.- $N1 = N1 - 1$; .

2.- Si $N1 = 0$, poner $TS1=inf$. Si no, Generar $DS1$, poner $TS1 = TM + DS1$

3.- Generar TP .

-Si $TP =$ servidor 2 (prob. p), $N2= N2+1$; Si $N2=1$, Generar $DS2$, poner $TS2 = TM + DS2$

3.- $SUMA = SUMA + NANT \times (TM-TANT)$

4.- $NANT=N1+N2$, $TANT=TM$

5.- Volver a Programa principal.

Servicio2

1.- $N2 = N2 - 1$.

2.- Si $N2 = 0$, poner $TS2= inf$. Si no, Generar $DS2$, poner $TS2 = TM + DS2$

3.- $SUMA = SUMA + NANT \times (TM-TANT)$

4.- $NANT=N1+N2$, $TANT=TM$

5.- Volver a Programa principal.

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Llegadas al servidor 1: Cte de 30 min.
- Distribución de los tiempos de servicio en la zona de inspección 1 (distribución trapezoidal).

$$f_1(x) = \begin{cases} \frac{x-20}{50} & 20 < x \leq 25 \\ 0.1 & 25 < x \leq 30 \\ \frac{35-x}{50} & 30 < x \leq 35 \\ 0 & \text{resto} \end{cases}$$

Valores para DS1 con la serie 1:

Serie 1: 0.5, 0.9, 0.3, 0.7, 0.9, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2

Modo de obtención de las variables aleatorias:

$$\begin{aligned} x &\in (20, 35); \\ c &\geq \max\{f(x) : x \in (20, 35)\}; \\ c &= 0,1; \\ \text{Genero} \\ u1, u2 \\ x &= 20 + 15 \times u1; \\ y &= 0,1 \times u2; \\ \text{Calculo } f(x) \\ \text{Si } y < f(x) &\rightarrow \text{Acepto } x; \end{aligned}$$

- Llegada de autobuses a la zona de reparación (discreta). Generación de TP.

Probabilidad de 0.3 que el autobús pase al servidor 2: Si $u < 0.3$, TP \rightarrow servidor2.

Probabilidad de 0.7 que el autobús se vaya definitivamente del sistema. Si $0.3 < u < 1$, TP \rightarrow Fuera

Usaremos la serie 2:

0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.2, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6

- Distribución de los tiempos de servicio en la zona de reparación (distribución Gamma de parámetros $p = 2$ y $a = 0,05$) truncada en 30 minutos:

Nota: Al ser el parámetro p natural, vamos a usar una Erlang(2, 0.05)

De la serie 3 calcularemos:
$$x = -\frac{\sum_{i=1}^p \ln u_i}{a};$$

Vamos a usar la serie 3:

0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5

TRAZA:

DS1: Tiempos de servicio del servidor 1 salen:

27,5 24,5 33,5 nulo 32 29 nulo 33,5 26

DS2: Tiempos de servicio de la ventanilla 2:

64,37 nulo 39,32 59,91 nulo nulo 34,29 31,21 37,94

TP: S2, Fuera, Fuera, Fuera, Fuera, S2, S2, Fuera, Fuera, , Fuera, Fuera Fuera, Fuera S2 S2 Fuera S2 fuera

Nº Evento	TM	Evento	N1	N2	TL	TS1	TS2	TP	SUMA
0	0	Inicio	0	0	30	∞	∞		0
1	30	Llegada	1		60	57.5	∞		0
2	57.5	Servicio1	0	1	60	∞	121.87	S2	27.5
3	60	Llegada	1	1	90	84.5	121.87		30
4	84.5	Servicio1	0	1	90	∞	121.87	Fuera	79
5	90	Llegada	1	1	120	123.5	121.87		84.5
6	120	Llegada	2	1	150	123.5	121.87		144.5
7	121.87	Servicio2	2	0	150	123.5	∞		150.11
8	123.5	Servicio1	1	0	150	155.5	∞	Fuera	153.37
9	150	Llegada	2	0	180	155.5	∞		179.87
10	155.5	Servicio1	1	0	180	184.5	∞	Fuera	190.87
11	180	Llegada	2	0	210	184.5	∞		215.37
12	184.5	Servicio1	1	0	210	218	∞	Fuera	224.37
13	210	Llegada	2	0	240	218	∞		249.87
14	218	Servicio1	1	1	240	244	257	S2	265.87
15	240	Llegada	2	1	270	244	257		309.87

Al cabo de las 4 horas el número total de autobuses es:

$N = N_1 + N_2 = 3$ autobuses en el taller de mantenimiento.

Al cabo de los 4 horas el número medio de autobuses es:

$SUMA/TM = 1,29$ autobuses cada minuto en el sistema.

RESULTADO DEL PROBLEMA 4

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

- Variables de estado: El objetivo de la simulación va a ser estudiar el número medio de pedidos sin atender que quedan al final del día; así pues, las variables de estado van a ser:
 $N(t) =$ Número de pedidos de ambos tipos que hay en el sistema.
- Eventos:
Los eventos que se pueden dar en nuestro sistema tal que produzcan un cambio en el mismo son:
 - Llegada de pedido.
 - Fabricación de pedido.
- Mecanismo de transición:
Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:
 - Llegada de pedido. $N \rightarrow N + 1$
 - Fabricación de pedido. $N_2 \rightarrow N - 1$
- Datos:
 - Distribución de tiempo entre pedidos. Es uniforme entre 3 y 5 horas.
 - Distribución de tiempo de fabricación de tipo de producto 1 (normal)
 - Distribución de tiempo de fabricación de tipo de producto 2 (triangular)
 - Tipo de pedido (discreta)
 - $TMAX =$ tiempo máximo de simulación. (40 horas).
- Otras variables:
 - **TM:** Reloj de simulación.
 - **DL:** Tiempo entre pedidos. (uniforme).
 - **DS1:** Tiempo de servicio tipo 1 = Normal.

- **DS2:** Tiempo de servicio tipo 2 = Normal.
- **TL:** Instante de la próxima llegada de cualquier tipo de pedido.
- **TS1:** Instante del próximo final de servicio de tipo 1.
- **TS2:** Instante del próximo final de servicio de tipo 2.
- **TP:** Tipo de pedido.

MODELO:

Programa principal:

0.- Inicialización

$N = 0$; $TM = 0$; $TS1 = \text{inf.}$; $TS2 = \text{inf.}$; $SUMA = 0$;

Generar DL ;

$TL = DL$;

1.- Actualización del reloj de simulación: $TM = \min(TL, TS1, TS2)$;

$TANT = TM$;

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL$. Llamar a subrutina: *Llegada a fabricaron.*

Si $TM = TS1$. Llamar a subrutina: *Servicio1.*

Si $TM = TS2$. Llamar a subrutina: *Servicio2.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar: $SUMA/TM$

Subrutinas

Llegada a fabricación

1.- $N = N + 1$;

2.- Generar DL ; $TL = TM + DL$;

3.- Si $N = 1$; Generar TP ;

Si $TP = 1$; Generar $DS1$ $TS1 = TM + DS1$; $TS2 = \text{inf.}$

Si $TP = 2$; Generar $DS2$ $TS2 = TM + DS2$; $TS1 = \text{inf.}$

4.- $SUMA = SUMA + (N - 1) \times (TM - TANT)$

5.- Volver a Programa principal.

Servicio1

1.- $N = N - 1$.

2.- Si $N = 0$, poner $TS1 = \text{inf.}$ $TS2 = \text{inf.}$

Si $N > 0$; Generar TP ;

Si $TP = 1$; Generar $DS1$ $TS1 = TM + DS1$; $TS2 = inf$.

Si $TP = 2$; Generar $DS2$ $TS2 = TM + DS2$; $TS1 = inf$.

3.- $SUMA = SUMA + (N - 1) \times (TM - TANT)$.

4.- Volver a Programa principal.

Servicio2

1.- $N = N - 1$.

2.- Si $N = 0$, poner $TS1 = inf$. $TS2 = inf$.

Si $N > 0$ Generar TP ;

Si $TP = 1$; Generar $DS1$ $TS1 = TM + DS1$; $TS2 = inf$.

Si $TP = 2$; Generar $DS2$ $TS2 = TM + DS2$; $TS1 = inf$;

3.- $SUMA = SUMA + (N - 1) \times (TM - TANT)$.

4.- Volver a Programa principal.

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Llegadas de pedidos: distribución uniforme de entre 3 y 5 horas.

Usamos la serie:

Serie 3: 0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5

Calculamos las variables aleatorias:

Dado $u \in (0,1)$; tenemos que $x = 3 + (5-3) \times u$;

Resultado: 3.2 3.8 4 4.2 4.4 3.4 3.2 4 4.4 4.8 4.6 4.2 4.8 3.4 3.6 4.4 3.6 4

- Los tiempos de fabricación de los pedidos de tipo 1: distribución normal de media 4 horas y desviación típica 1.

Vamos a usar como números aleatorios:

Serie 1: 0.5, 0.9, 0.3, 0.7, 0.9, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2

Dado $u \in (0,1)$; tenemos que $x = 1 \times u + 4$;

Resultado:

4.5, 4.9, 4.3, 4.7, 4.9, 4.2, 4.1, 4.6, 4.8, 4.3, 4.6, 4.1, 4.8, 4.7, 4.9, 4.1, 4.4, 4.2

- Los tiempos de fabricación de los pedidos de tipo 2: distribución triangular simétrica entre 2 y 4 horas.

Vamos a usar como números aleatorios:

Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.2, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6

Dado $u_1, u_2 \in (0,1)$; tenemos el triángulo de base (2,4) y altura (0,1):

Método:

$$x \in (2, 4);$$

$$c \geq \max\{f(x) : x \in (2, 4)\};$$

$$c = 1;$$

$$\text{Genero } u_1, u_2$$

$$x = 2 + 2 \times u_1;$$

$$y = 1 \times u_2;$$

Calculo $f(x)$

Si $y < f(x) \rightarrow$ Acepto x ;

Resultado: No acepto; No acepto; No acepto; 3,4; 2,4; 3,6; No acepto; No acepto; 3,4:

- Lo tipos de pedido se distribuyen según una discreta; vamos a usar:

Serie 4: 0.2, 0.7, 0.5, 0.3, 0.1, 0.9, 0.5, 0.6, 0.2, 0.8, 0.7, 0.4, 0.5, 0.6, 0.9, 0.3, 0.1, 0.4

Probabilidad de 0.35 que sea de tipo 1: Si $u < 0.35$, $x \rightarrow 1$

Probabilidad de 0.65 que sea de tipo 2.: Si $0.35 < u < 1$, $x \rightarrow 2$

Valores para TP según la serie 4: 1 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 1 2

TRAZA:

DL: Tiempo de llegadas al sistema:

3.2 3.8 4 4.2 4.4 3.4 3.2 4 4.4 4.8 4.6 4.2 4.8 3.4 3.6 4.4 3.6 4

DS1: Tiempos de fabricación de pedido 1:

4.5, 4.9, 4.3,

4.7, 4.9, 4.2, 4.1, 4.6, 4.8, 4.3, 4.6, 4.1, 4.8, 4.7, 4.9, 4.1, 4.4, 4.2

DS2: Tiempos de fabricación de pedido 2:

No acepto; No acepto; No acepto; 3,4; 2,4; 3,6; No acepto; No acepto; 3,4:

TP: tipo de pedido: 1 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 1 2

Nº evento	Reloj TM	Tipo evento	N	TL	TS1	TS2	TP	SUMA
0	0	Inicio	0	3.2	∞	∞		0
1	3.2	Llegada	1	7	7.7	∞	1	0
2	7	Llegada	2	11	7.7	∞		3.8
3	7.7	Servicio1	1	11	∞	11.1	2	3.8
4	11	Llegada	2	15.2	∞	11.1		7.1
5	11.1	Servicio2	1	15.2	∞	13.5	2	7.1
6	13.5	Servicio2	0	15.2	∞	∞		7.1
7	15.2	Llegada	1	19.6	∞	18.8	2	7.1
8	18.8	Servicio2	0	19.6	∞	∞		7.1
9	19.6	Llegada	1	23	∞	23	2	7.1
10	23	Llegada	2	26.4	∞	23		10.5
11	23	Servicio2	1	26.4	27.9	∞	1	10.5
12	26.4	Llegada	2	29.6	27.9	∞		13.9
13	27.9	Servicio1	1	29.6	32.2	∞	1	13.9
14	29.6	Llegada	2	33.6	32.2	∞		15.6
15	32.2	Servicio1	1	33.6	∞	35.6	2	15.6
16	33.6	Llegada	2	38	∞	35.6		17
17	35.6	Servicio2	1	38	∞	38	2	17
18	38	Llegada	2	42.8	∞	38		19.4
19	38	Servicio2	1	42.8	∞	41.6		19.4

Al cabo de estas 40 horas se han dejado sin atender 1 pedido. Se han dado servicio a una media de $19.4/40 = 0.485$ productos cada hora.

RESULTADO DEL PROBLEMA 9

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

➤ Variables de estado:

El objetivo de la simulación va a ser estudiar el factor de utilización de la impresora.

$NES(t)$ = Número de trabajos que esperan para ser procesados por el servidor

$NEI(t)$ = Número de trabajos que esperan en el buffer para ser impresos o están siendo impresos.

$N(t)$ = Número de trabajos impresos.

➤ Eventos:

Los eventos que se pueden dar en nuestro sistema produciendo un cambio en el mismo son:

- Llegada de un trabajo al servidor.
- Servicio de un trabajo en el servidor = Llegada de cliente al buffer ó directamente a impresión.
- Impresión de un trabajo (lo que equivale a irse definitivamente del sistema).

➤ Mecanismo de transición:

Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:

- Llegada de un cliente al servidor 1 : $NES \rightarrow NES+1$.
- Servicio de un trabajo en el servidor =Llegada de un cliente al servidor 2
 $NES(t) \rightarrow NES(t)-1$;
 $NEI(t) \rightarrow NEI(t)+1$.
- Servicio de un cliente en el servidor2: $NEI(t) \rightarrow NEI(t)-1$.
 $N(t) \rightarrow N(t) + 1$.

➤ Datos:

- Distribución de tiempos entre llegadas al servidor (exponencial de media 5 minutos)
- Distribución de los tiempos de servicio del servidor (trapezoidal)
- Distribución de los tiempos de impresión de la impresora ($G(x)$)
- TMAX = tiempo máximo de simulación. (30 minutos).

➤ Otras variables:

- **TM:** Reloj de simulación.

- **TU:** tiempo que la impresora ha estado trabajando.
- **DL:** Tiempo entre llegadas = Exponencial.
- **DS1:** Tiempo de servicio servidor = Trapezoidal.
- **DS2:** Tiempo de servicio impresora = $G(x)$.
- **TL:** Instante de la próxima llegada.
- **TS1:** Instante del próximo final de servicio en servidor.
- **TS2:** Instante del próximo final de servicio en impresora.

MODELO:

Programa principal:

0.- Inicialización

NES = 0; NEI = 0; N = 0; TM = 0; TS1 = inf.; TS2 = inf.; TU = 0

Generar DL; TL = DL

1.- Actualización del reloj de simulación: $TM = \min(TL, TS1, TS2)$;

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL$. Llamar a subrutina: *Llegada a servidor.*

Si $TM = TS1$. Llamar a subrutina: *Servicio.*

Si $TM = TS2$. Llamar a subrutina: *Impresión.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar: NES, NEI, N, factor de utilización = TU/TM ;

Subrutinas

Llegada a servidor

1.- NES = NES + 1;

2.- *Generar DL*, poner $TL = TM + DL$.

Si NES = 1, *Generar DS1*, poner $TS1 = TM + DS1$;

3.- Volver a Programa principal.

Servicio

1.- NES = NES - 1; NEI = NEI + 1.

2.- Si NES = 0, poner $TS1 = \text{inf.}$ Si no, *Generar DS1*, poner $TS1 = TM + DS1$;

Si $NEI = 1$, Generar $DS2$, poner $TS2 = TM + DS2$; $TU = TU + DS2$.

3.- Volver a Programa principal.

Impresión

1.- $NEI = NEI - 1$. $N = N + 1$;

2.- Si $NEI = 0$, poner $TS2 = inf$. Si no, Generar $DS2$, poner $TS2 = TM + DS2$;

$TU = TU + DS2$.

3.- Volver a Programa principal.

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Los tiempos de llegada al sistema: distribución exponencial de media 5

Se emplea el método de la transformada inversa; mediante los números aleatorios uniforme

$$u \in [0,1) \text{ obtenemos DL: } x = -\frac{\ln(u)}{1/5}$$

Usamos serie 1: 0.5, 0.9, 0.3, 0.7, 0.2, 0.1, 0.6, 0.8, 0.3, 0.6, 0.1, 0.8, 0.7, 0.9, 0.1, 0.4, 0.2, 0.9

Valores para DL con la serie 1:

3.47 0.53 6.02 1.78 8.05 11.51 2.55 1.12 6.02 2.55 11.51 1.12 1.78 0.53 11.51 4.58 8.05 0.53

- Los tiempos de servicio del servidor: distribución trapezoidal de función de densidad:

$$f(x) = \begin{cases} \frac{4}{3}x & 0 < x < 0.5 \\ \frac{2}{3} & 0.5 < x < 1.25 \\ \frac{2}{3}(2.25 - x) & 1.25 < x < 2.25 \\ 0 & \text{resto} \end{cases}$$

Usaremos la serie 2:

Serie 2: 0.2, 0.8, 0.1, 0.5, 0.9, 0.6, 0.7, 0.1, 0.3, 0.6, 0.8, 0.3, 0.4, 0.5, 0.2, 0.9, 0.7, 0.6

Modo de obtención de las variables aleatorias:

$x \in (0.2, 25)$;
 $c \geq \max\{f(x) : x \in (0.2, 25)\}$;
 $c = 2 / 3$;
Genero u_1, u_2
 $x = 2,25 \times u_1$;
 $y = 2 / 3 \times u_2$;
Calculo $f(x)$
Si $y < f(x) \rightarrow$ *Acepto* x ;

Valores para DS1: 0.45 no acepto no acepto 1.575 0.675 1.8 0.9 0.45 1.575

-Los tiempos de servicio de la impresora DS2: función de distribución $G(x)$:

$$G(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{12} & 1 < x < 10 \\ \frac{x+5}{20} & 10 < x < 15 \\ 1 & x > 15 \end{cases}$$

Usamos el método de la transformada inversa:

Si $u < 0.75 \rightarrow x = 12 \otimes u + 1$

Si $u > 0.75 \rightarrow x = 20 \otimes u - 5$

Usamos la serie 3:

Serie 3: 0.1, 0.4, 0.8, 0.6, 0.7, 0.2, 0.1, 0.5, 0.7, 0.9, 0.8, 0.6, 0.9, 0.2, 0.3, 0.7, 0.3, 0.5

Resultados para DS2:

4.6 5.8 11 8.2 9.4 3.4 2.2 7 9.4 13 11 8.2 13 3.4 4.6 9.4 4.6 7

TRAZA:

DL: 3.47 0.53 6.02 1.78 8.05 11.51 2.55 1.12 6.02 2.55 11.51 1.12 1.78 0.53 11.51 4.58
8.05 0.53

DS1: 0.45 no acepto no acepto 1.575 0.675 1.8 0.9 0.45 1.575

DS2: 4.6 5.8 11 8.2 9.4 3.4 2.2 7 9.4 13 11 8.2 13 3.4 4.6 9.4 4.6 7

Nº evento	Reloj TM	Tipo evento	NES	NEI	TL	TS1	TS2	TU	N
0	0	Inicio	0	0	3.47	∞	∞	0	0
1	3.47	Llegada	1		4	3.92	∞	0	0
2	3.92	Servicio	0	1	4	∞	8.52	4.6	0
3	4	Llegada	1	1	10.02	5.75	8.52	4.6	0
4	5.75	Servicio	0	2	10.02	∞	8.52	4.6	0
5	8.52	Impresión	0	1	10.02	∞	14.32	10.4	1
6	10.02	Llegada	1	1	11.8	10.7	14.32	10.4	1
7	10.7	Servicio	0	2	11.8	12.5	14.32	10.4	1
8	11.8	llegada	1	2	19.85	12.5	14.32	10.4	1
9	12.5	Servicio	0	3	19.85	∞	14.32	10.4	1
10	14.32	Impresión	0	2	19.85	∞	25.32	21.4	2
11	19.85	Llegada	1	2	31.36	20.75	25.32	21.4	2
12	20.75	Servicio	0	2	31.36	∞	25.32	21.4	2
13	25.32	Impresión	0	1	31.36	∞	33.52	29.6	3

El factor de utilización resulta de dividir el número de trabajos impresos entre el tiempo de utilización de la máquina en los 30 minutos simulados.

Factor de utilización = $3/29.6 = 0.101$.

RESULTADO DEL PROBLEMA 10

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

➤ Variables de estado:

El objetivo de la simulación va a ser estudiar el coste por hora del primer robot:

$N1(t)$ = Número de maquinas que hay en el taller de reparación.

R: Variable que me dice si el robot esta funcionando ($R = 0$) o esta arreglándose ($R = 1$).

➤ Eventos:

Los eventos que se pueden dar en nuestro sistema tal que produzcan un cambio en el mismo son:

- Llegada de una máquina al taller.

- Reparación de una maquina = Salida de máquina del taller.
- Inicio avería de un robot.
- Fin reparación del robot

➤ Mecanismo de transición:

Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:

- Llegada de una máquina al taller : $N1 \rightarrow N1+1$.
- Reparación de una maquina = Salida de máquina del taller.

$$N1 \rightarrow N1-1;$$

- Avería de un robot. $N1$ se mantiene. $R=0$
- Fin de reparación de robot. $R=1$

➤ Datos:

- Distribución de tiempos entre llegadas de máquinas al taller (Poisson de parámetro 5)
- Distribución de los tiempos de reparación (exponencial)
- Distribución de los tiempos de avería del robot: $f(x)$
- Tiempo de duración de reparación robot 15 min.
- TMAX = tiempo máximo de simulación. (120 minutos).

➤ Otras variables:

- **TM:** Reloj de simulación.
- **DL:** Tiempo entre llegadas = Poisson
- **DS1:** Tiempo de reparación = Exponencial.
- **DAR:** Tiempo entre averías de robot. $F(x)$.
- **TL:** Instante de la próxima llegada.
- **TS1:** Instante del próximo final de reparación.
- **TAR:** Instante del próximo final de avería de robot.
- **TFR:** instante del próximo final de reparación de robot
- **SUMA:** contador acumulado suma de áreas de taller en el sistema por tiempo de permanencia.
- **TANT:** variable auxiliar.
- **NANT:** variable auxiliar.
- **A:** Variable auxiliar que me dice cuanto tiempo trabaja el robot.
- **T(t) :** Tiempo que el robot esta funcionando.

MODELO:

Programa principal:

0.- Inicialización

$N1 = 0; N2 = 0; TM = 0; TS1 = \text{inf.}; TFR = \text{inf.}; NNAT = 0; TANT = 0; R=0 ; A = 0 ;$

Generar DL; TL = DL

Generar DAR; TAR = DAR;

1.- Actualización del reloj de simulación: $TM = \min (TL, TS1, TAR, TFR);$

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL$. Llamar a subrutina: *Llegada a taller.*

Si $TM = TS1$. Llamar a subrutina: *Reparación.*

Si $TM = TAR$. Llamar a subrutina: *Avería.*

Si $TM = TFR$. Llamar a subrutina: *Fin Avería.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar:

Coste = $1000 \times 2 + 2000 \times A/60 + 8000 \times SUMA/120$

Subrutinas

Llegada a taller

1.- Si $R = 0$

1.- $N1 = N1 + 1;$

Si $N1 = 1$, Generar DS1, poner $TS1 = TM + DS1, A = DS1 + A;$

$NANT=N1, TANT=TM$

Volver a Programa principal.

2.- Si $R = 1$

$N1 = N1 + 1$

Si $N1 = 1$

Generar DS1, $TS1 = TFR + DS1, A = DS1 + A;$

3. Generar DL, poner $TL = TM + DL.$

$SUMA = SUMA + NANT \times (TM - TANT)$

$NANT=N1, TANT=TM$

4. Volver a Programa principal.

Reparación

- 1.- $N1 = N1 - 1$;
- 2.- Si $N1 = 0$, poner $TS1 = \text{inf.}$ Si no, *Generar* $DS1$, poner $TS1 = TM + DS1$; $A = A + DS1$;
- 3.- $SUMA = SUMA + NANT \times (TM - TANT)$
- 4.- $NANT = N1$, $TANT = TM$
- 5.- Volver a Programa principal.

Avería

$$R = 1$$

- 1.- $TFR = TAR + 15$;
- Si $N \geq 1$
- $$TS1 = TS1 + 15$$
- 2.- $SUMA = SUMA + NANT \times (TM - TANT)$
 - 3.- $NANT = N1$, $TANT = TM$
 - 4.- Volver a Programa principal.

Fin avería

$$R = 0$$

- 1.- $TFR = \text{infinito}$;
- Generar DAR ; $TAR = DAR + TM$;
- 2.- Volver a programa principal

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Llegadas al taller: distribución exponencial de media $1/5 * 60 = 12$ minutos
- $$X = -\ln(u) * 12$$

Valores para DL con la serie 1:

8,31 1,26 14,44 4,28 19,31 27,63 6,12 14,44 6,12 27,63 2,67 4,28 1,26 27,63 10,99 19,31
1,26 2,67

- Los tiempos de reparación: distribución exponencial. El robot 1 arregla. 6 máquinas por hora: Exponencial de media 10 minutos.

$$X = -\text{Ln}(u) \times 10;$$

Valores para DS1 con la serie 2:

16,09 2,23 22,07 6,93 1,05 5,10 23,02 12,03 5,10 2,23 12,03 9,16 6,93 16,09 1,05 3,56 5,10 3,56

- Los tiempos de servicio del servidor: distribución de función de densidad.:

$$f(x) = \begin{cases} \frac{32}{3}(x - 0.75) & 0.75 \leq x \leq 1 \\ \frac{16}{3}(1.5 - x) & 1 \leq x \leq 1.5 \\ 0 & \text{resto} \end{cases}$$

Usaremos la serie 3:

Serie 3: 0.9, 0.1, 0.8, 0.6, 0.2, 0.5, 0.7, 0.4, 0.1, 0.6, 0.3, 0.7, 0.3, 0.1, 0.1, 0.4, 0.3, 0.1

Modo de obtención de las variables aleatorias:

$$x \in (0, 75.1, 5);$$

$$c \geq \max\{f(x) : x \in (0, 75.1, 5)\};$$

$$c = 2,66$$

$$\text{Genero } u1, u2$$

$$x = 0,75 + 0,5 \times u1;$$

$$y = 2,66 \times u2;$$

$$\text{Calculo } f(x)$$

$$\text{Si } y < f(x) \rightarrow \text{Acepto } x;$$

Valores para DAR (en minutos): 72 No acepto 69 63 51 60 66 57 48

TRAZA:

DL: Tiempo de llegada al taller: 8,31 1,26 14,44 4,28 19,31 27,63 6,12 14,44 6,12 27,63 2,67 4,28 1,26 27,63 10,99 19,31 1,26 2,67

DS1: Tiempos de reparación del robot 1: 16,09 2,23 22,07 6,93 1,05 5,10 23,02 12,03 5,10 2,23 12,03 9,16 6,93 16,09 1,05 3,56 5,10 3,56

Valores para DAR (en minutos): Tiempos de avería robot. 72 No acepto 69 63 51 60 66 57 48

Nº evento	Reloj TM	Tipo evento	N1	N2	A	TL	TS1	TAR	TFR	R	SUMA
0	0	Inicio	0	0	0	8.31	∞	72	∞	0	0
1	8.31	Llegada	1	0	16.09	9.57	24.4	72	∞	0	0
2	9.57	Llegada	2	0	16.09	24.01	24.4	72	∞	0	1.26
3	24.01	Llegada	3	0	16.09	28.29	24.4	72	∞	0	30.14
4	24.4	Reparación	2	0	18.32	28.29	26.63	72	∞	0	31.31
5	26.63	Reparación	1	0	40.39	28.29	48.7	72	∞	0	35.77
6	28.29	Llegada	2	0	40.39	47.6	48.7	72	∞	0	37.42
7	47.6	Llegada	3	0	40.39	75.23	48.7	72	∞	0	76.04
8	48.7	Reparación	2	0	47.32	75.23	55.63	72	∞	0	79.34
9	55.63	Reparación	1	0	48.37	75.23	56.68	72	∞	0	93.2
10	56.68	Reparación	0	0	48.37	75.23	∞	72	∞	0	94.25
11	72	Avería	0	0	48.37	75.23	∞	72	87	1	94.25
12	75.23	Llegada	1	0	53.47	77.9	92.1		87	1	94.25
13	77.9	Llegada	2	0	53.47	82.18	92.1		87	1	96.92
14	82.18	Llegada	3	0	53.47	83.44	92.1		87	1	105.48
15	83.44	Llegada	4	0	53.47	111.07	92.1		87	1	109.26
16	87	Fin avería	4	0	53.47	111.07	92.1	141	∞	0	109.26
17	92.1	Reparación	3	0	76.49	111.07	115.02	141	∞	0	129.66
18	111.07	Llegada	4	0	76.49	122.07	115.02	141	∞	0	186.57
19	115.92	Reparación	3	0	88.52	122.07	126.8	141	∞	0	205.97

Coste = $1000 \times 2 + 2000 \times A/60 + 8000 \times \text{SUMA}/120 = 18679$ euros en dos horas. 9339 Euros / hora.

RESULTADOS DEL PROBLEMA 13

En primer lugar se presentará el modelo con todos los elementos que incluye y en segundo lugar se presentará la traza, junto con los métodos utilizados para generar las variables aleatorias que se describen.

ELEMENTOS DEL MODELO:

➤ Variables de estado:

El objetivo de la simulación va a ser estudiar el número de clientes que hay en cada una de las ventanillas, así pues, las variables de estado van a ser:

- N_1 = Número de clientes que hay en el servidor 1 (ventanilla para comprobar que la declaración es correcta).
- N_2 = Número de clientes que hay en el servidor 2 (ventanilla para pagar o recibir el importe de la declaración).

Nota: El número total de clientes en nuestro sistema va a ser N_1+N_2 .

➤ Eventos:

Los eventos que se pueden dar en nuestro sistema tal que produzcan un cambio en el mismo son:

- Llegada de un cliente al servidor 1.
- Servicio de un cliente en el servidor 1 = Llegada de cliente al servidor 2
- Servicio de un cliente en el servidor 2 (lo que equivale a irse definitivamente de la agencia o sistema).

➤ Mecanismo de transición:

Los cambios que se producen en el estado del sistema cuando se produce uno de los eventos anteriormente descritos son:

- Llegada de un cliente al servidor 1 : $N_1 \leftarrow N_1+1$.
- Servicio de un cliente del servidor 1 =Llegada de un cliente al servidor 2
 $N_1 \rightarrow N_1-1$;
 $N_2 \rightarrow N_2+1$.
- Servicio de un cliente en el servidor2: $N_2 \leftarrow N_2-1$.

➤ Datos:

- Distribución de tiempos entre llegadas a la ventanilla 1 (exponencial de media 1 min.)
- Distribución de los tiempos de servicio de la ventanilla 1 (discreta)
- Distribución de los tiempos de servicio de la ventanilla 2 (discreta)
- TMAX = tiempo máximo de simulación. (10 minutos).

➤ Otras variables:

- **TM:** Reloj de simulación.
- **DL:** Tiempo entre llegadas = Exponencial.
- **DS1:** Tiempo de servicio ventanilla1 = Discreta.
- **DS2:** Tiempo de servicio ventanilla2 = Discreta.
- **TL:** Instante de la próxima llegada.
- **TS1:** Instante del próximo final de servicio en ventanilla1.
- **TS2:** Instante del próximo final de servicio en ventanilla2.

MODELO:

Programa principal:

0.- Inicialización

$N1 = 0; N2 = 0; TM = 0; TS1 = \text{inf.}; TS2 = \text{inf.};$

Generar DL; TL = DL

1.- Actualización del reloj de simulación: $TM = \min (TL, TS1, TS2);$

2.- Identificación de evento y llamada a subrutinas de evento.

Si $TM = TL$. Llamar a subrutina: *Llegada a servidor 1.*

Si $TM = TS1$. Llamar a subrutina: *Servicio1.*

Si $TM = TS2$. Llamar a subrutina: *Servicio2.*

3.- Regla de parada.

Si $TM < TMAX$, ir a 1. Si no, parar: $N1, N2$.

Subrutinas

Llegada a servidor1

1.- $N1 = N1 + 1;$

2.- *Generar DL*, poner $TL = TM + DL$.

Si $N1 = 1$, *Generar DS1*, poner $TS1 = TM + DS1;$

3.- Volver a Programa principal.

Servicio1

1.- $N1 = N1 - 1; N2 = N2 + 1.$

2.- Si $N1 = 0$, poner $TS1 = \text{inf.}$ Si no, *Generar DS1*, poner $TS1 = TM + DS1$

Si $N2 = 1$, *Generar DS2*, poner $TS2 = TM + DS2;$

3.- Volver a Programa principal.

Servicio2

1.- $N2 = N2 - 1$.

2.- Si $N2 = 0$, poner $TS2 = \text{inf}$. Si no, Generar $DS2$, poner $TS2 = TM + DS2$;

3.- Volver a Programa principal.

MÉTODOS DE GENERACIÓN DE VARIABLES ALEATORIAS:

- Llegadas a la ventanilla 1: distribución exponencial de media 1 minuto: $x = -\ln(u)$;

Valores para DL con la serie 1:

0.27, 2.81, 1.83, 0.23, 0.61, 0.52, 1.77, 0.12, 0.69, 0.07, 1.1, 0.44, 0.67, 0.31, 1.46, 0.47.

- Los tiempos de servicio de la ventanilla 1: distribución discreta.

Probabilidad de 0.2 que sean 0.6 minutos: Si $u < 0.2$, $x \leftarrow 0.6$

Probabilidad de 0.5 que sean 0.7 minutos: Si $0.2 < u < 0.7$, $x \leftarrow 0.7$

Probabilidad de 0.3 que sean 0.8 minutos: Si $u > 0.7$, $x \leftarrow 0.8$

Valores para $DS1$ según la serie 2:

0.6, 0.6, 0.7, 0.7, 0.8, 0.7, 0.6, 0.7, 0.6, 0.7, 0.7, 0.8, 0.6, 0.8, 0.7, 0.7

- Los tiempos de servicio de la ventanilla 2: distribución discreta.

Probabilidad de 0.1 que sean 0.8 minutos: Si $u < 0.1$, $x \leftarrow 0.8$

Probabilidad de 0.7 que sean 0.9 minutos: Si $0.1 < u < 0.8$, $x \leftarrow 0.9$

Probabilidad de 0.2 que sea 1 minuto: Si $u > 0.8$, $x \leftarrow 1$

Valores para $DS2$ según la serie 3:

0.9, 0.9, 0.9, 1, 0.8, 0.9, 0.9, 0.9, 1, 0.9, 0.8, 0.9, 0.9, 0.9, 1, 0.9

TRAZA:

DL: Tiempo de llegada a la ventanilla 1:

0.27, 2.81, 1.83, 0.23, 0.61, 0.52, 1.77, 0.12, 0.69, 0.07, 1.1, 0.44, 0.67, 0.31

$DS1$: Tiempos de servicio de la ventanilla 1:

0.6, 0.6, 0.7, 0.7, 0.8, 0.7, 0.6, 0.7, 0.6, 0.7, 0.7, 0.8, 0.6, 0.8, 0.7, 0.7

$DS2$: Tiempos de servicio de la ventanilla 2:

0.9, 0.9, 0.9, 1, 0.8, 0.9, 0.9, 0.9, 1, 0.9, 0.8, 0.9, 0.9, 0.9, 1, 0.9

Nº evento	Reloj TM	Tipo evento	N1	N2	TL	TS1	TS2
0	0	Inicio	0	0	0.27	∞	∞
1	0.27	Llegada	1	0	3.08	0.87	∞
2	0.87	Servicio1	0	1	3.08	∞	1.77
3	1.77	Servicio2	0	0	3.08	∞	∞
4	3.08	Llegada	1	0	4.91	3.68	∞
5	3.68	Servicio1	0	1	4.91	∞	4.58
6	4.58	Servicio2	0	0	4.91	∞	∞
7	4.91	Llegada	1	0	5.14	5.61	∞
8	5.14	Llegada	2	0	5.75	5.61	∞
9	5.61	Servicio1	1	1	5.75	6.31	6.51
10	5.75	Llegada	2	1	6.27	6.31	6.51
11	6.27	Llegada	3	1	8.04	6.31	6.51
12	6.31	Servicio1	2	2	8.04	7.11	6.51
13	6.51	Servicio2	2	1	8.04	7.11	7.51
14	7.11	Servicio1	1	2	8.04	7.81	7.51
15	7.51	Servicio2	1	1	8.04	7.81	8.31
16	7.81	Servicio1	0	2	8.04	∞	8.31
17	8.04	Llegada	1	2	8.16	8.64	8.31
18	8.16	Llegada	2	2	8.85	8.64	8.31
19	8.31	Servicio2	2	1	8.85	8.64	9.21
20	8.64	Servicio1	1	2	8.85	9.34	9.21
21	8.85	Llegada	2	2	8.92	9.34	9.21
22	8.92	Llegada	3	2	10.02	9.34	9.21
23	9.21	Servicio2	3	1	10.02	9.34	10.11
24	9.34	Servicio1	2	2	10.02	9.94	10.11
25	9.94	Servicio1	1	3	10.02	10.64	10.11
26	10.02	Parada	1	3			

IX Prácticas propuestas para finanzas

Estas prácticas están propuestas como guía en el inicio a la simulación aplicada a finanzas, con orden creciente de complejidad, sobre todo del instrumento financiero sobre el que se proponen. Pueden programarse en cualquier lenguaje, herramienta u hoja de cálculo (Matlab, Excel...).

PRÁCTICA 1: GENERADOR CONGRUENCIAL Y NORMAL

1. Programar un generador congruencial ($m = 2^{31} - 1$, $a=16807$, $b=0$)
2. Generar dos series de normales mediante Box-Müller ($x = \sqrt{-2 \ln u_1} \cos(2\pi u_2)$,
 $y = \sqrt{-2 \ln u_1} \sin(2\pi u_2)$)
3. Hacer el histograma, y dar la media, varianza, coeficientes de asimetría y curtosis, el VaR y la correlación entre ambas series.

PRÁCTICA 2: VARIABLES ANTITÉTICAS Y ESTIMACIÓN

1. Generar las antitéticas de las muestras anteriores.
2. Estimar la correlación entre las muestras originales y sus antitéticas. Estimar VaR, media y varianza de las muestras antitéticas y de las formadas por las muestras originales y las antitéticas

PRÁCTICA 3: MOVIMIENTO BROWNIANO

1. Generar un movimiento browniano unidimensional para $T=100$, con $\Delta t = 1$ según la expresión $X \leftarrow X + \sqrt{\Delta t} N(0,1)$. Representar gráficamente
2. Generar un movimiento browniano bidimensional para $T=100$, con $\Delta t = 1$ según la expresión $X \leftarrow X + \sqrt{\Delta t} N(0,1), Y \leftarrow Y + \sqrt{\Delta t} N(0,1)$. Representar gráficamente
3. Generar una muestra de valores de un movimiento browniano en $T=100$, con $\Delta t = 1$ según las dos opciones siguientes y comparar los resultados de la estimación de media y varianza
 - a. $X \leftarrow X + \sqrt{\Delta t} N(0,1)$
 - b. $Z = \sqrt{T} N(0,1)$

PRÁCTICA 4: VALORACIÓN OPCIÓN CALL EUROPEA VAINILLA

Obtener el valor de una opción call europea vainilla con vencimiento T. Dar la estimación del valor si $V_{call} = e^{-r_t} E_{riesgoneutro}[\text{payoff}] = (1+r)^{-t} E_{riesgoneutro}[\text{payoff}]$ y la precisión (tened en cuenta que la varianza de la estimación del valor es $S^2_{V_{call}} = (e^{-r_t})^2 S^2_{\text{payoff}}$).

1. Obtener el valor final del subyacente mediante la expresión en el plazo final

$$S(t) = S(0)e^{\left(r_c - d_c - \frac{1}{2}\sigma^2\right)t + \sigma\sqrt{t}N(0,1)}$$

2. Obtener el valor final del subyacente mediante simulación dinámica con $\Delta t = 1/12$ y la

$$\text{expresión } S \leftarrow S e^{\left(r_c - d_c - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}N(0,1)}$$

3. Obtener el valor final del subyacente mediante simulación dinámica con $\Delta t = 1/12$ y el mecanismo de transición

$$S \leftarrow S + S(r_{\Delta t} - d_{\Delta t}) + \sigma S \sqrt{\Delta t} N(0,1) = S e^{(r_c - d_c)\Delta t} + \sigma S \sqrt{\Delta t} N(0,1) = S(1+r-d)^{\Delta t} + \sigma S \sqrt{\Delta t} N(0,1)$$

4. Comparar los resultados anteriores con la fórmula de Black-Scholes (

$$V_{call}(S, t) = S e^{-d_c} \Phi(d_1) - K e^{-r_t} \Phi(d_2), \text{ donde}$$

$$d_1 = \frac{\ln(S/K) + (r_c - d_c + 1/2\sigma^2)t}{\sigma\sqrt{t}} \text{ y } d_2 = d_1 - \sigma\sqrt{t}$$

Utilizar datos anuales: $r = 3\%$, $\text{Strike} = 14$, $\text{Plazo } T = 1, 2 \text{ y } 3 \text{ años}$, $\sigma = 0,3$; $S_0 = 14$, d (divid)=0%

PRÁCTICA 5: VALORACIÓN OPCIÓN CALL EUROPEA VAINILLA NO BROWNIANA (T DE STUDENT)

Obtener mediante simulación el valor de una opción call europea vainilla con dividendos para plazos $t=1, 2, \dots, T$, pero cuyos incrementos se rigen según una distribución T de Student. Dar la estimación del valor y la precisión y comparar con el resultado de la práctica anterior. Utilizar los mismos datos de la práctica anterior, y $n = 6$. (Nota: recordar que la varianza de una T de

Student con n grados de libertad es $\frac{n}{n-2}$, con lo que al sustituir la normal por la T de Student,

ha de hacerse con ésta tipificada). Utilizar para la evolución del subyacente

$$S \leftarrow S e^{(r_c - d_c)\Delta t} + \sigma S \sqrt{\Delta t} \frac{T_n}{\sqrt{n/(n-2)}}.$$

PRÁCTICA 6: VALORACIÓN OPCIÓN CALL EUROPEA VAINILLA BROWNIANA CON SALTOS

Obtener mediante simulación el valor de una opción call europea vainilla en la que el subyacente evoluciona según un proceso de Wiener y un proceso de Poisson. Es decir, suponer que sobre el proceso de Wiener actúa un proceso de Poisson de parámetro $\lambda = 2$ (anual) para saltos en la valoración, y que esos saltos se distribuyen según $X_S = \begin{cases} 10\%S & 1/2 \\ -15\%S & 1/2 \end{cases}$. Dar la estimación del valor y la precisión, y comparar con el resultado de las prácticas anteriores.

PRÁCTICA 7: VALORACIÓN OPCIÓN PUT EUROPEA Y AMERICANA: ALGORITMO BINOMIAL

Obtener mediante el algoritmo binomial el valor de una opción put europea y una opción put americana, con los datos de la práctica 4, y comparar el resultado de la estimación del valor de la opción europea con el obtenido mediante simulación (usar la relación de paridad sobre el valor obtenido en el ejercicio 4).

PRÁCTICA 8: VALORACIÓN OPCIÓN CALL ASIÁTICA

Obtener mediante simulación el valor de una opción call asiática browniana con dividendos y vencimiento T. Dar la estimación del valor y la precisión. Utilizar los mismos datos que en la práctica 4, y comparar el resultado con los obtenidos en esa práctica.

PRÁCTICA 9: VALORACIÓN OPCIÓN PUT BERMUDAS

Obtener el valor de una opción put bermudas browniana y con opción de ejercicio trimestral. Utilizar los mismos datos que en la práctica 4, y comparar con los resultados de la práctica 7.

PRÁCTICA 10: VALORACIÓN ESTRATEGIA BUTTERFLY

Obtener mediante simulación el valor de una estrategia butterfly formada por la compra de una opción call europea con strike K, la compra de otra opción europea con strike $K' > K$, y la venta de dos call europeas con strike K'' tal que $K < K'' < K'$, todas sobre el mismo subyacente. Aplicar suponiendo los datos del apartado 4 con $K=14$, $K'=16$ y $K''=15$.

PRÁCTICA 11: VALOR Y RIESGO DE UNA CARTERA DE OPCIONES

Obtener mediante simulación el valor de una cartera formada por opciones call y put europeas sobre diversos subyacentes relacionados entre sí, y comparar con el valor supuesto que son independientes. Obtener el VaR de los payoff de la cartera a los 3 años. Suponer la siguiente composición de la cartera (una unidad de cada opción) y los siguientes datos: $r = 5\%$
Correlación $\rho_{12} = -0,7$

Opción 1: Subyacente S_1 , call Strike = 15, Plazo $T = 3$ años, sigma = 0,3, $S_1(0) = 14$

Opción 2: Subyacente S_1 , put Strike = 14, Plazo $T = 3$ años, sigma = 0,3, $S_1(0) = 14$

Opción 3: Subyacente S_2 , call Strike = 11, Plazo $T = 3$ años, sigma = 0,4, $S_2(0) = 10$