

Simulación de sistemas discretos

Noviembre de 2006

Álvaro García Sánchez
Miguel Ortega Mier

1.	Presentación.....	4
1.1.	Introducción	4
1.2.	Sistemas, modelos y simulación.....	4
1.3.	Necesidad de la simulación.....	8
1.4.	Campos de aplicación	9
1.5.	Fases en un estudio de simulación.....	10
1.6.	Ventajas de la simulación	14
1.7.	Inconvenientes de la simulación	15
1.8.	Principales errores cometidos en la simulación	15
1.9.	Simulación de sistemas discretos	16
2.	Repaso de estadística	21
2.1.	Introducción	21
2.2.	VARIABLES ALEATORIAS.....	22
2.3.	Media y varianza	25
2.4.	VARIABLES ALEATORIAS MÁS COMÚNMENTE UTILIZADAS.....	26
2.5.	Estimación de parámetros	35
2.6.	Ajuste de datos. Test de la χ^2	37
2.7.	Intervalos de confianza.....	41
2.8.	Comparación de alternativas	43
3.	Formulación del problema y modelo conceptual	50
3.1.	Recogida, análisis y generación de datos de entrada	51
4.	Construcción del modelo. Verificación, validación y credibilidad	59
4.1.	Entornos de simulación.	60
4.2.	Verificación.....	61
4.3.	Validación.....	65
5.	Ejecución de un modelo de simulación. Análisis de resultados.....	68

5.1.	Tipos de simulación	68
5.2.	Análisis en simulación limitada	70
5.3.	Análisis en simulación ilimitada con régimen permanente	73
6.	Referencias y bibliografía	79

1. Presentación

1.1. Introducción

El presente documento trata sobre las técnicas utilizadas para imitar o *simular* el funcionamiento de distintos tipos de instalaciones o procesos. A la instalación o proceso que se pretende estudiar se le denomina *sistema* y para poderlo analizar se realiza una serie de supuestos sobre su funcionamiento. Estos supuestos, que normalmente se expresan mediante relaciones matemáticas o relaciones lógicas, constituyen un *modelo* del sistema. Este modelo se utiliza para comprender y prever el comportamiento del sistema real.

Si las relaciones matemáticas o lógicas que comprende el modelo son sencillas, entonces será posible utilizar un procedimiento analítico para obtener una solución o respuesta exacta sobre las características de interés del sistema analizado. No obstante, si las relaciones son complejas, puede ocurrir que no se pueda evaluar analíticamente el problema. En este caso, será necesario acudir a la *simulación* del sistema, evaluando *numéricamente* el modelo y analizando los datos obtenidos para *estimar* las características de dicho sistema.

1.2. Sistemas, modelos y simulación

Un *sistema* se puede definir como un conjunto de elementos unidos por relaciones de interacción o interdependencia. En el ámbito de los sistemas productivos estos elementos normalmente tienen un objetivo común.

Los elementos que forman parte del sistema vienen condicionados por el objetivo del estudio que se pretende realizar, ya que un sistema definido para un estudio determinado puede ser una parte de un sistema más amplio

definido para otro estudio particular. Por ejemplo, si se quiere determinar cuál es el número más adecuado de operarios y máquinas en la sección de mecanizado de una empresa que tiene una determinada cartera de pedidos, estos elementos serán los que formen parte del sistema a analizar, mientras que, si lo que se desea es estudiar la capacidad productiva de la empresa, los elementos mencionados anteriormente sólo serán una parte del sistema. A ellos habrá que añadir montaje, embalaje, almacenaje, etc.

Se pueden realizar las siguientes definiciones:

- *Atributo*: propiedad de un elemento del sistema.
- *Actividad*: todo proceso que provoque un cambio en el sistema.

El *estado* del sistema en un instante de tiempo determinado se puede definir como la descripción de todos los elementos, atributos y actividades en dicho instante. Por ejemplo, el estado de una oficina bancaria en un instante se podría definir mediante el número de cajeros en él, el número de clientes, el instante de llegada de cada cliente y el tipo de operación que desea realizar cada uno. Este conjunto constituiría las *variables de estado* del sistema.

Tipos de sistemas

Evidentemente, las características del sistema real que se desea estudiar van a condicionar el tipo de simulación que se va a desarrollar. Por lo tanto, conviene hacer una clasificación de los sistemas de acuerdo con los aspectos que van a condicionar su análisis posterior. Así, es útil realizar una clasificación de los sistemas atendiendo a tres aspectos fundamentales:

- **Sistemas estáticos y sistemas dinámicos.** Un sistema se considera estático cuando sus variables de estado no cambian a lo largo del tiempo, es decir, cuando el tiempo no juega ningún papel en sus propiedades. Por el contrario, en un sistema dinámico los valores que toman todas o algunas de sus variables de acción evolucionan a lo largo del tiempo.
- **Sistemas deterministas y sistemas estocásticos.** Si un sistema no tiene ningún componente de carácter estocástico (es decir, aleatorio) se

considera determinista. En este caso, el comportamiento del sistema está determinado una vez que se hayan definido las condiciones iniciales y las relaciones que existen entre sus componentes. Por el contrario, un sistema no determinista o estocástico tiene algún elemento que se comporta de forma aleatoria, de forma que no está predeterminado comportamiento en función de las condiciones iniciales y de las relaciones entre sus componentes. En este caso, el sistema sólo se podrá estudiar en términos probabilistas, consiguiendo, en el mejor de los casos, conocer sus respuestas posibles con sus probabilidades asociadas.

- **Sistemas continuos y sistemas discretos.** En un sistema continuo las variables de estado cambian de forma continua a lo largo del tiempo, mientras que en uno discreto cambian instantáneamente de valor en ciertos instantes de tiempo. En un sistema de una cierta complejidad puede ocurrir que existan simultáneamente variables de estado continuas y discretas. En este caso, dependiendo de la predominancia de una y otras y del objetivo del estudio que se pretende realizar, se considerará el sistema como perteneciente a uno de los dos tipos.

Tipos de modelos

Para estudiar un sistema, la forma más inmediata sería experimentar sobre él. Sin embargo, esto puede ser desaconsejable, e incluso imposible, por diversos motivos:

- Puede ocurrir que el sistema no exista y lo que se pretenda sea su diseño.
- Puede ser imposible experimentar con el sistema real porque no se dispone de ningún control sobre dicho sistema; por ejemplo, si se desea estudiar un sistema financiero, bursátil,...
- Puede ser económicamente inviable la experimentación sobre el sistema real.
- La experimentación sobre el sistema real puede conllevar unos plazos de tiempo muy dilatados. Es el caso, por ejemplo, de ciertos sistemas sociales o biológicos.

En cualquiera de los casos anteriores se hace necesaria la construcción de un modelo del sistema que refleje con la fidelidad adecuada las características destacadas del sistema a analizar y la experimentación sobre dicho modelo. Si se realiza correctamente la construcción del modelo y el diseño de los experimentos, los resultados obtenidos permitirán inferir cuál sería el comportamiento del sistema a analizar en determinadas condiciones.

La figura 1 muestra un esquema que muestra las diferentes formas de simulación que se pueden utilizar para analizar un sistema.

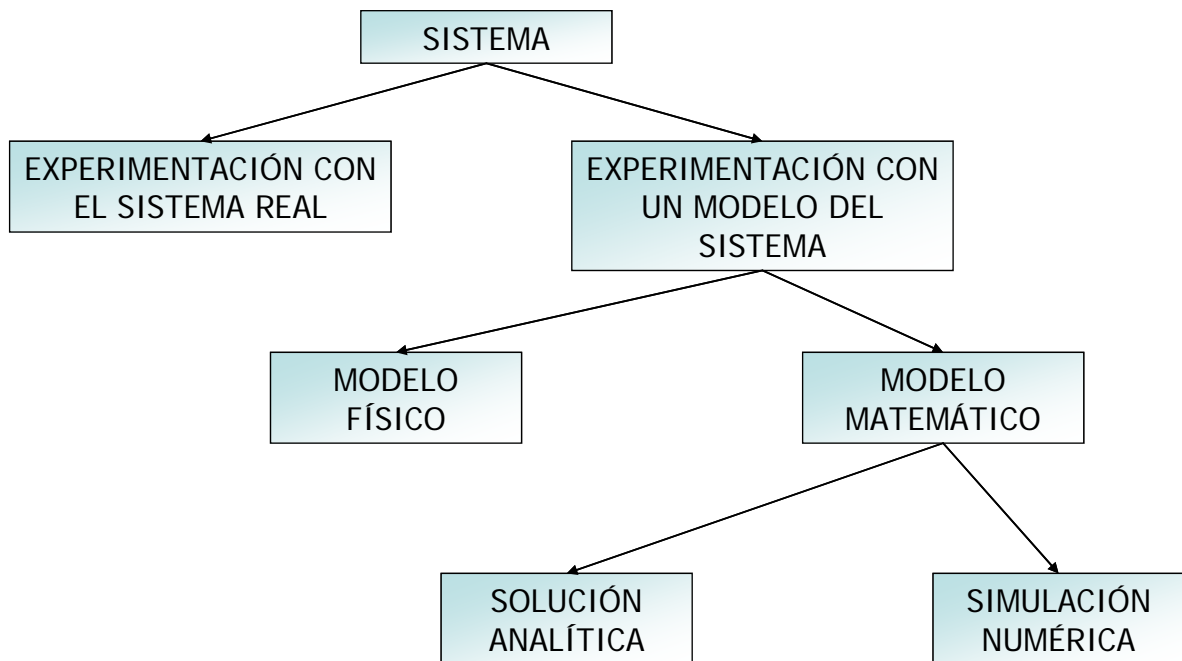


Fig. 1. Tipos de modelos

Los modelos físicos están formados por una estructura material que tiene unas características, en cuanto al objeto del estudio, similares a las del sistema real. Ejemplos de modelos físicos pueden ser las maquetas a escala y también los modelos analógicos que, sin tener la misma estructura física que el sistema real, tienen un comportamiento similar con respecto a algunas variables de estado. Por ejemplo, para estudiar una red de distribución de agua en una ciudad, se puede construir un circuito eléctrico

con la misma estructura y establecer analogías entre la intensidad eléctrica del modelo y el caudal, la resistencia eléctrica de los distintos tramos y la pérdida de presión. De este modo, se puede prever el comportamiento de la red de distribución realizando experimentos e, incluso, modificando el modelo eléctrico sin necesidad de actuar sobre el sistema real, lo que implicaría, evidentemente, unos costes muy elevados y un deterioro en el servicio prestado a los clientes.

Un modelo matemático representa el sistema por medio de relaciones lógicas y cuantitativas entre sus variables de estado. Tanto el valor de las variables como sus relaciones se pueden modificar para estudiar cómo reacciona el modelo y, por lo tanto, cómo reaccionaría el sistema real ante dichos cambios. Aunque en casos puntuales se han desarrollado modelos físicos, para el estudio de los sistemas productivos en la gran mayoría de las ocasiones se utilizan modelos matemáticos.

Una vez que se ha construido un modelo matemático que representa el sistema a estudiar, se debe analizar cómo utilizar este modelo para resolver las preguntas planteadas sobre el sistema. Si el modelo matemático es suficientemente sencillo se podrá resolver analíticamente, obteniendo una solución exacta a dichas preguntas. Si el modelo es excesivamente complejo o inabordable, habrá que recurrir a su simulación, que consiste en proporcionar una serie de valores a determinadas variables de estado y calcular cuál es el valor resultante para el resto de las variables. De este modo, se obtiene una representación o muestra de las posibles respuestas de modelo (y, por lo tanto, del sistema que representa) ante distintas condiciones de partida.

1.3. Necesidad de la simulación

Ya se ha indicado anteriormente que se recurre a la simulación cuando el modelo matemático que representa el sistema a estudiar es excesivamente complejo o resulta inabordable por no estar desarrollados

métodos analíticos para su resolución. La fuente de complejidad puede tener básicamente dos causas:

- En los sistemas continuos es frecuente que unas variables de estado representen la tasa o velocidad de cambio de otras variables de estado. La formulación matemática de estos modelos lleva a la aparición de ecuaciones diferenciales que indican las relaciones anteriormente mencionadas. Si el sistema tiene una cierta complejidad, puede ocurrir que las ecuaciones diferenciales sean no lineales y, por lo tanto, de difícil o imposible resolución analítica.
- En los sistemas discretos pueden aparecer fenómenos aleatorios que sólo se pueden representar en términos probabilistas. En este caso, la formulación matemática del modelo contiene relaciones en las que aparecen funciones de distribución o de densidad de probabilidad, que dificultan o impiden su resolución analítica.

Como ya se ha indicado, la catalogación de un sistema como continuo o discreto depende del objetivo del estudio y de las variables de estado predominantes. Esto quiere decir que un mismo sistema puede tener ciertas variables de estado continuas y otras discretas. Por lo tanto, no es infrecuente encontrar modelos en los que coexisten ecuaciones diferenciales complejas con variables aleatorias, lo que, evidentemente, complica aún más la resolución analítica.

1.4. Campos de aplicación

Como la simulación de modelos de sistemas reales exige la resolución numérica de los sistemas de ecuaciones planteados y, como consecuencia, la realización de un número muy elevado de cálculos, la simulación requiere necesariamente el empleo de ordenadores. El aumento de la capacidad de los ordenadores que se ha producido en los últimos años, así como el desarrollo de distintos paquetes de software diseñados específicamente para la simulación, cada vez más potentes y de utilización más sencilla, han hecho

que la simulación se haya generalizado para el estudio de sistemas de muy distinta naturaleza. Sin hacer una recopilación exhaustiva, se puede destacar la utilización cada vez más extendida de la simulación en el estudio de:

- Sistemas de espera.
- Tráfico de comunicaciones: correos, teléfonos, redes informáticas...
- Diseño de instalaciones, talleres, líneas de montaje...
- Determinación de reglas de programación de la producción.
- Diseño de plantillas, asignación de trabajadores a puestos de trabajo...
- Localización de instalaciones (almacenes, vehículos, equipos de mantenimiento...)
- Análisis de proyectos.
- Reglas de gestión de inventarios.
- Análisis de inversiones.

1.5. Fases en un estudio de simulación

En la figura 2 se indican las fases fundamentales de que consta un estudio de simulación:

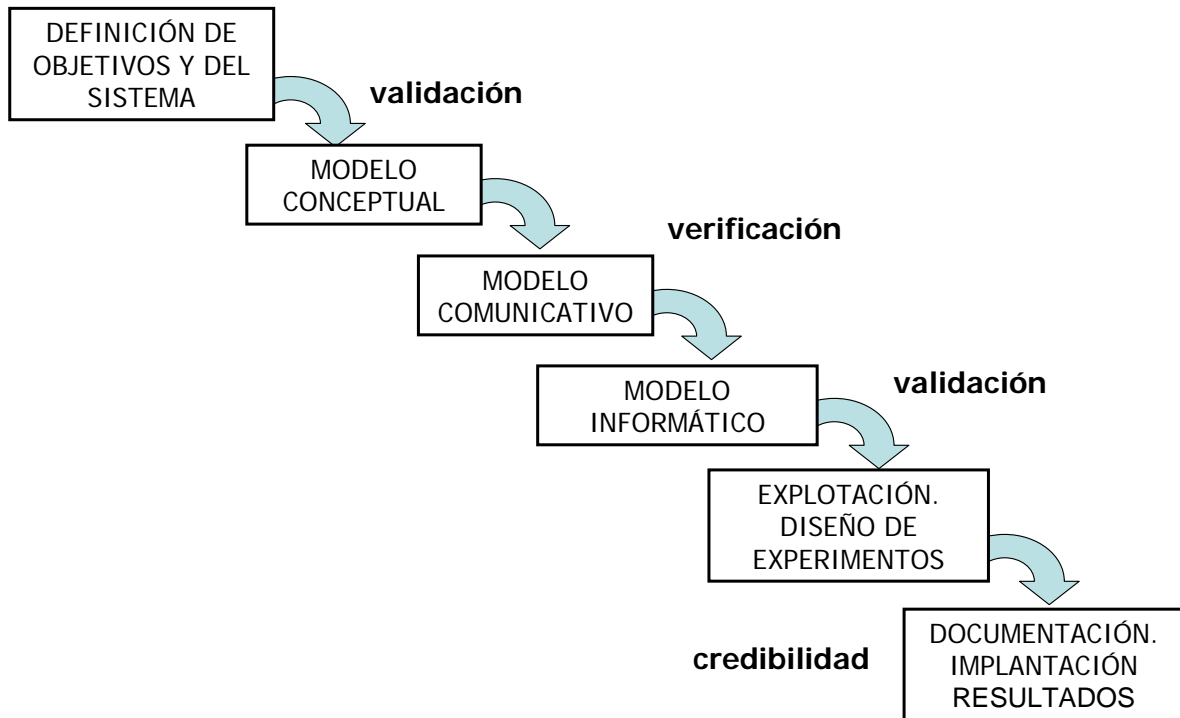


Fig. 2. Etapas de un estudio de simulación

Definición de objetivos y del sistema. En primer lugar, se deben especificar claramente los objetivos que se pretenden alcanzar con el estudio de simulación. Se deben traducir los objetivos cualitativos a términos cuantitativos, establecer las preguntas que deben ser contestadas, las hipótesis a contrastar, y los efectos a estimar. También es necesario introducir los criterios de evaluación de los resultados y realizar una estimación de los medios humanos y materiales para llevar a cabo dicho estudio. Es necesario, además, definir los elementos que van a formar parte del sistema objeto de estudio; muy probablemente, el sistema sea un subsistema de otro más amplio con el que interactúa.

Elaboración del modelo conceptual. El modelo conceptual es un modelo lógico y matemático del sistema real, diseñado de acuerdo con los objetivos que se pretenden alcanzar con el estudio. En la construcción del modelo es aconsejable encontrar un equilibrio entre la sencillez del propio modelo y el realismo con que representa al sistema real. Muchos autores aconsejan comenzar con un modelo relativamente sencillo, a partir del cual elaborar modelos de complejidad creciente. Un modelo debe tener

únicamente el grado necesario de detalle que refleje la esencia del funcionamiento del sistema desde el punto de vista del propósito para el que se utiliza dicho modelo. En la mayoría de los casos no es necesario que exista una correspondencia biunívoca entre los elementos del modelo y los del sistema. En esta fase es necesario estimar los valores de las constantes y los parámetros, determinar los valores iniciales de las diferentes variables y, si es posible, recoger datos históricos para la validación del modelo.

Validación. Una vez definido el modelo conceptual, será necesario validarlo, es decir, comprobar si refleja fielmente las características del sistema que representa. En esta fase pueden ser de gran ayuda las intervenciones y opiniones de personas que conozcan con suficiente profundidad el sistema.

Elaboración del modelo comunicativo. Lo más común es que los responsables del sistema y los responsables últimos del estudio de simulación sean diferentes de los programadores que después realizarán el modelo informático. Por eso es necesario elaborar algún tipo de modelo que permita que la comunicación entre los diseñadores y los programadores sea eficaz y eficiente. Los diagramas de flujo, en los que se representan los diferentes eventos son especialmente útiles.

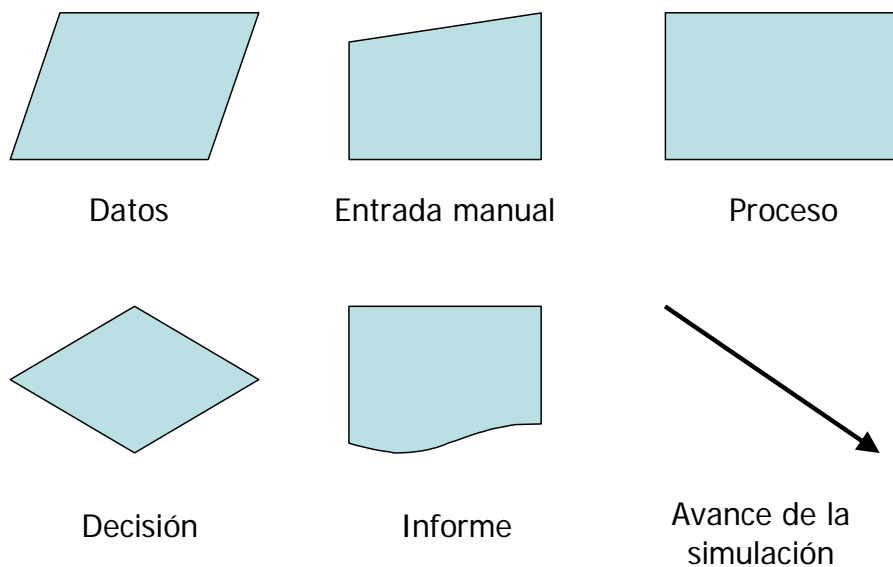


Fig. 3. Modelo comunicativo. Elementos más frecuentemente utilizados.

Construcción y verificación de modelo informático. Una vez construido y validado el modelo es el momento de seleccionar el lenguaje de programación que se va a utilizar para su realización. En función de las características del modelo se puede optar por un lenguaje de propósito general, como Fortran, Pascal o C, o bien un lenguaje desarrollado especialmente para la simulación, como GPSS, Simgen, Slam o Siman. Por otra parte, en los últimos años se han desarrollado paquetes de software con capacidades gráficas de animación, especialmente útiles en el diseño de sistemas productivos y logísticos, ya que permiten visualizar a lo largo del tiempo los movimientos y estados de máquinas, piezas, vehículos, transportadores, etc. Entre estos paquetes se encuentran Cinema, Witness, Taylor o Arena.

Validación. El modelo anterior se debe validar mediante la ejecución de una serie de experimentos piloto, en los que los resultados obtenidos coincidan con los previsibles ante determinadas condiciones iniciales. Por otra parte, si el sistema modelado es similar a alguno ya existente, se puede contrastar el funcionamiento del modelo con el del sistema real.

Explotación y diseño de experimentos. De acuerdo con los objetivos de la simulación, se deben definir los experimentos a realizar. Para cada uno de ellos es necesario determinar las condiciones iniciales, la longitud de la simulación, el número de repeticiones y los resultados que se deben registrar. Para analizar los resultados de los distintos experimentos se utilizan técnicas estadísticas. Los análisis típicos pueden ser el establecimiento de intervalos de confianza para los valores obtenidos de ciertas variables de estado o la comparación y determinación del mejor de los resultados obtenidos en la simulación de varias alternativas.

Elaboración de la documentación e implantación de los resultados. Ya que los modelos de simulación, a menudo, se utilizan para más de una aplicación, es importante no sólo el programa de ordenador, sino

también las hipótesis bajo las cuales se ha construido el modelo. Es de destacar que cuanto mejor documentado y más verosímil sea un modelo de simulación más probabilidad tendrá de ser utilizado, se dirá que es más 'creíble'.

1.6. Ventajas de la simulación

Ya se ha comentado previamente que la simulación es una técnica cada vez más utilizada en el estudio de sistemas complejos. Entre los argumentos a favor de la utilización de la simulación se encuentran los siguientes:

- La mayoría de los sistemas complejos reales con elementos estocásticos no se pueden describir con suficiente precisión mediante un modelo matemático que se pueda resolver analíticamente. Por lo tanto, con frecuencia la simulación es el único método posible de estudio de dichos sistemas.
- La simulación permite estimar el comportamiento de un sistema existente bajo un conjunto previsto de condiciones operativas.
- Mediante la simulación se pueden comparar diseños alternativos (o políticas de operación alternativas para un determinado diseño) para especificar cuál es el que cumple de forma más adecuada con los objetivos formulados.
- En la simulación se puede tener un control mucho mejor sobre las condiciones del experimento que si se realizase sobre el propio sistema.
- La simulación permite estudiar un sistema cuya evolución es muy dilatada en el tiempo (por ejemplo, un sistema económico) en un periodo de tiempo reducido. Alternativamente, también permite estudiar de forma detallada la evolución de un sistema en un corto periodo de tiempo.

1.7. Inconvenientes de la simulación

La simulación no sólo ofrece ventajas, sino que puede plantear algunos inconvenientes. Entre ellos se encuentran:

- Cada ejecución de un modelo estocástico de simulación da como resultado únicamente una estimación de las características o comportamiento del modelo para un conjunto particular de valores de las variables de entrada. Por lo tanto, no bastará con la ejecución del modelo una sola vez, sino que habrá que realizar una serie de repeticiones para obtener una *muestra representativa* del funcionamiento del sistema. En consecuencia, la decisión inherente a la formulación del problema deberá tomarse en base a dicha muestra y sin el conocimiento de todas las posibles respuestas del modelo. Esto no ocurre si se puede resolver el modelo analíticamente, ya que en este caso, se conocerán todas las respuestas del modelo. Por consiguiente, si se dispone de un modelo analítico válido que se pueda desarrollar sin una gran dificultad, será preferible a un modelo de simulación.
- Los modelos de simulación, por regla general, consumen una cantidad elevada de recursos técnicos y humanos durante un tiempo prolongado.
- La gran cantidad de información que proporcionan los modelos de simulación, así como la capacidad de persuasión que tienen algunos paquetes con animación gráfica, hacen que, a menudo, se confíe en exceso en los resultados que proporcionan. Si un modelo de simulación no proporciona una representación “válida” del sistema real, la información que suministra puede no ser válida o, incluso, puede conducir a la toma de decisiones erróneas.

1.8. Principales errores cometidos en la simulación

La experiencia demuestra que existe una serie de errores en los que frecuentemente se incurre al realizar un estudio de simulación. Entre ellos se pueden destacar:

- No definir correctamente los objetivos del estudio.
- Fijar un nivel de detalle inadecuado en el modelo.
- Tratar el estudio de simulación como si fuese principalmente un ejercicio complicado de programación.
- Utilizar un software de simulación comercial que no pueda reflejar de forma adecuada la lógica de funcionamiento del modelo.
- Utilizar de forma inadecuada la animación.
- Determinar de forma inadecuada las fuentes de aleatoriedad en el sistema real.
- Emplear funciones de distribución de probabilidad distintas a las correspondientes a los fenómenos reales que se quieren simular.
- Analizar los datos resultantes de la simulación considerando, en las fórmulas estadísticas utilizadas, que todos los valores son independientes.
- Realizar un número de repeticiones menor del necesario y considerar significativos los resultados obtenidos.

1.9. Simulación de sistemas discretos

Es evidente que los sistemas productivos evolucionan a lo largo del tiempo y, por lo tanto, debe tenerse en cuenta su naturaleza dinámica. Por otra parte, aunque determinados sistemas productivos pueden ser considerados como continuos (por ejemplo, reactores químicos), la mayoría de los sistemas productivos tienen características de sistemas discretos, ya que los cambios de estado (recepción de materias primas, inicio y finalización de la fabricación de lotes, entradas y salidas del almacén,...) se producen en instantes de tiempo determinados y separados entre sí. Por este motivo, esta documentación se centra en la simulación de sistemas dinámicos y discretos.

Mecanismos de avance del tiempo

Debido a la naturaleza dinámica de los modelos representativos de los sistemas productivos y logísticos, será necesario, durante la simulación, llevar un registro del valor actual del tiempo simulado mientras se desarrolla el experimento de simulación, así como un mecanismo para hacer avanzar este tiempo de un valor a otro. A la variable que determina el tiempo actual en un experimento de simulación se le suele denominar *reloj*. Cuando se utiliza un lenguaje de ordenador de propósito general para realizar la simulación, no se define explícitamente las unidades (horas, minutos...) con las que va a trabajar el reloj, sino que la unidad de tiempo corresponde a la unidad utilizada para las variables de entrada. Por otra parte, no suele existir ninguna correspondencia entre la unidad de tiempo elegida y el tiempo de ordenador necesario para ejecutar un experimento de simulación.

Históricamente, se han utilizado dos mecanismos distintos para hacer avanzar el tiempo en la simulación: *intervalos de tiempo variables* e *intervalos de tiempo fijos*.

Intervalos de tiempo variables

Es el método más utilizado. Consiste en inicializar el reloj a cero y determinar los instantes de ocurrencia de los sucesos de cada tipo más cercanos en el tiempo. A continuación, se incrementa el tiempo de reloj para hacerlo coincidir con el instante de ocurrencia del *suceso más próximo* de entre todos los sucesos futuros. En este instante se actualizan las variables de estado del sistema (ya que se ha producido un suceso y, por lo tanto, se ha producido un cambio de estado), se registran los valores de interés para el experimento y se determinan los nuevos instantes de ocurrencia de los sucesos futuros (si es necesario). Una vez realizadas las operaciones anteriores, se vuelve a incrementar el tiempo hasta el instante de ocurrencia del suceso más próximo y se repite el proceso. Los pasos anteriores se

repite hasta que se alcance alguna condición determinada o se llegue al tiempo de simulación fijado.

En resumen, el tiempo avanza siempre desde un instante en que se ha producido un suceso hasta el próximo instante en que se va a producir un nuevo suceso. Debido a ello, los diferentes incrementos de tiempo no tienen por qué tomar el mismo valor.

Es esquema de la figura 4 muestra la organización de un experimento de simulación realizado mediante el enfoque de intervalos de tiempos variables.

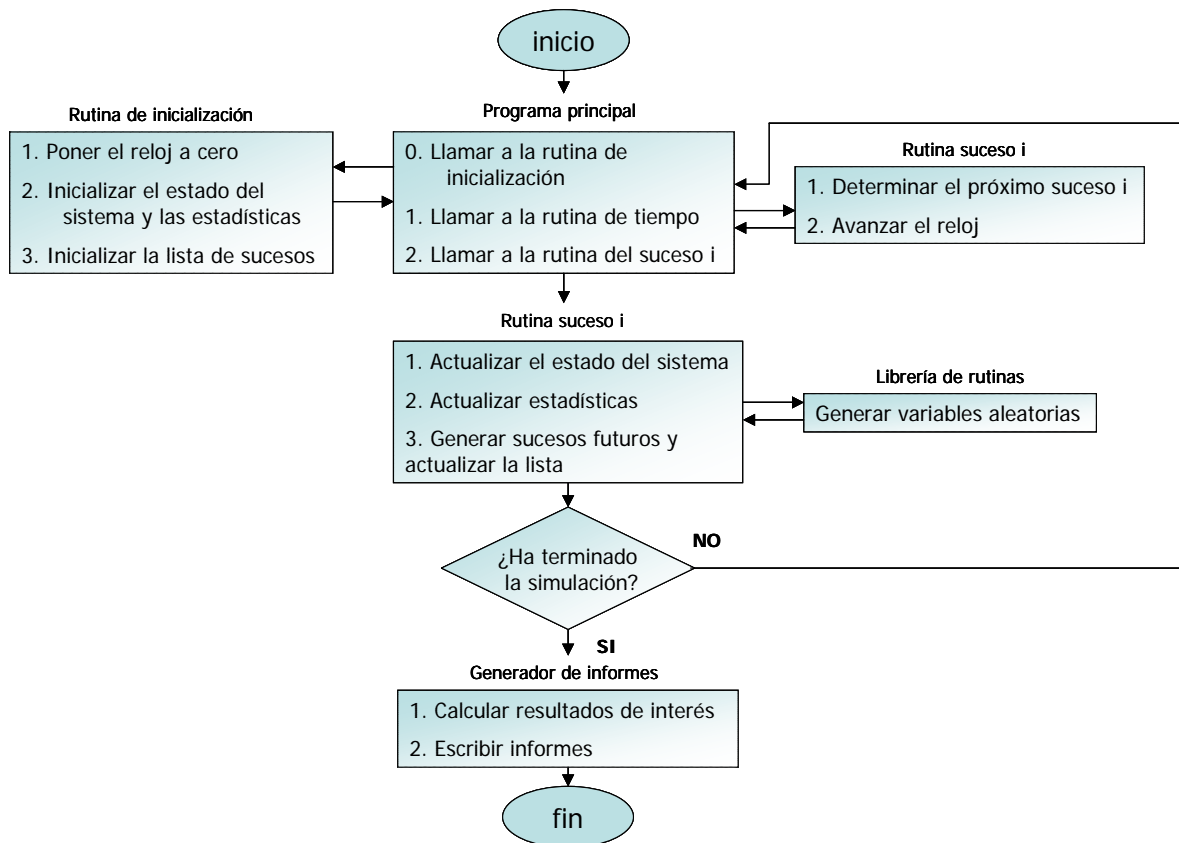


Fig. 4. Flujograma de la simulación con intervalos de tiempo variables

Aunque, como ya se ha indicado, la simulación de sistemas dinámicos en tiempo discreto se ha utilizado en una gran cantidad de aplicaciones, la mayoría de los modelos tienen unos componentes en común, que son los siguientes:

- *Estado del sistema*: conjunto de variables de estado necesarias para describir el sistema en un instante determinado de tiempo.
- *Reloj*: variable que recoge el valor actual del tiempo simulado.
- *Lista de sucesos*: lista que recoge el siguiente instante en que va a ocurrir cada tipo de suceso.
- *Estadísticas*: variables utilizadas para almacenar información estadística sobre el funcionamiento del modelo.
- *Rutina de inicialización*: subprograma para inicializar el modelo de simulación en el instante cero.
- *Rutina de tiempo*: subprograma que determina el siguiente suceso de la lista de sucesos y avanza el reloj al instante en que este suceso se produce.
- *Rutina de sucesos*: subprograma que actualiza el estado del sistema cuando ocurre un tipo particular de suceso (existe una rutina de sucesos para cada tipo de sucesos).
- *Librería de rutinas*: conjunto de programas utilizados para generar observaciones aleatorias de las distribuciones de probabilidad correspondientes a los sucesos aleatorios del modelo.
- *Generador de informes*: subprograma que, a partir de las estadísticas, calcula las estimaciones de las medidas de funcionamiento del modelo y produce un informe cuando ha terminado la simulación.
- *Programa principal*: subprograma que llama a la rutina de tiempo para determinar la ocurrencia del próximo suceso y transfiere el control a la rutina de sucesos correspondiente para actualizar adecuadamente el estado del sistema. También chequea la terminación del experimento y llama al generador de informes cuando el experimento ha terminado.

Intervalos de tiempo fijos

De acuerdo con este enfoque, el reloj avanza en incrementos de igual valor. Después de cada actualización del reloj, hay que realizar un chequeo para determinar si ha ocurrido algún suceso durante el intervalo Δt

inmediatamente anterior. Si han ocurrido uno o más sucesos en dicho intervalo, se considera que se han producido al final del intervalo y el estado del sistema (y las estadísticas) se deben actualizar de acuerdo con este supuesto.

Este procedimiento tiene dos desventajas principales; una de ellas es debida a los errores que se cometen al considerar que los sucesos se producen al final del periodo, y la otra es que si se producen dos o más sucesos en un Δt , ha de tomarse la decisión del orden en el cual se han producido si la realidad no permite considerarlos simultáneos. Estos dos problemas se pueden paliar si se reduce la longitud del incremento elemental de tiempo considerado. Sin embargo, una disminución de Δt conlleva inevitablemente un aumento del tiempo de ejecución del modelo en el ordenador.

Debido a las consideraciones anteriores, el procedimiento de incrementos de tiempo fijos no se suele emplear en modelos en los cuales los intervalos de tiempo entre sucesos pueden variar de forma considerable.

Alguna frase para cerrar este capítulo e introducir el siguiente

2. Repaso de estadística

2.1. Introducción

Como se ha señalado en el capítulo anterior, los fenómenos que son típicamente objeto de estudio mediante simulación son de carácter estocástico, por lo que la simulación está muy ligada a la estadística. En particular, y como se indica en la figura 5, a lo largo del desarrollo de un estudio de simulación, la estadística aparece de la siguiente manera.

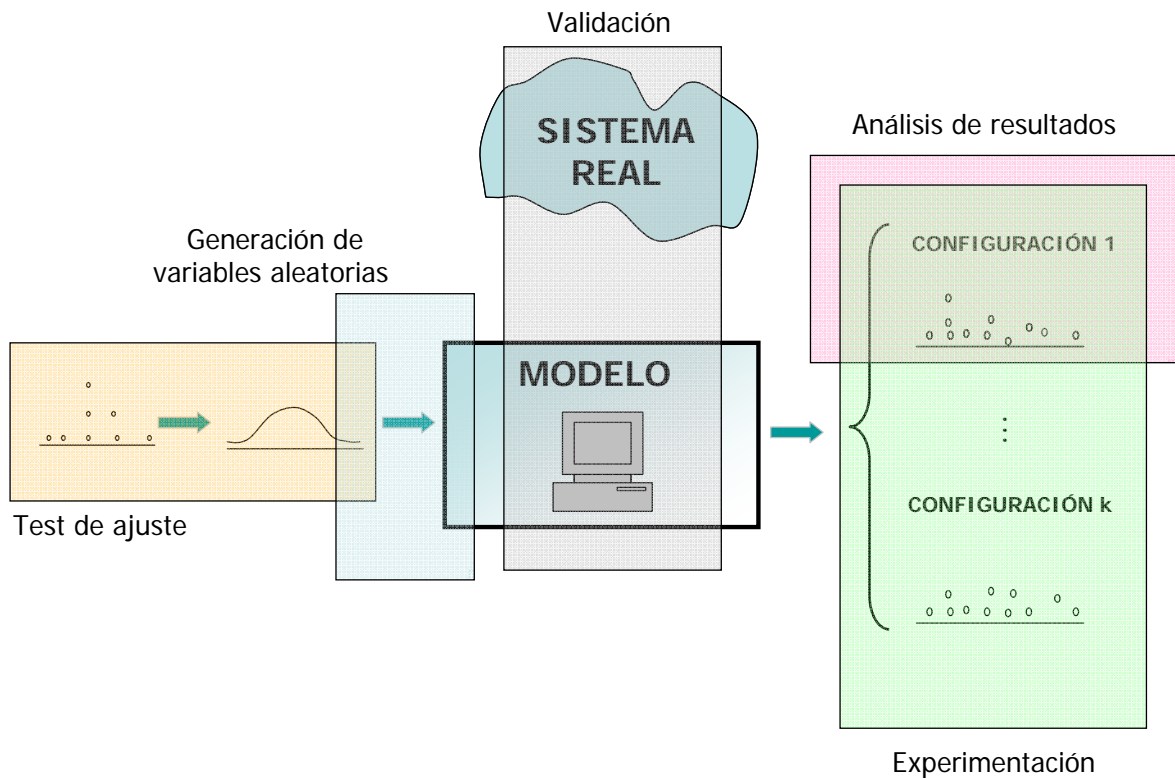


Fig. 5. La simulación y la estadística

- En primer lugar, se deben tratar de forma adecuada los datos históricos correspondientes a las variables de entrada para poder caracterizar de forma adecuada su comportamiento.

- Una vez hecho lo anterior, es necesario generar valores de las variables de entrada conforme a la caracterización del comportamiento anterior.
- Se debe estudiar de forma adecuada el valor de las variables de salida para no extraer conclusiones erróneas o no representativas del comportamiento del sistema.
- Se debe poder comparar de forma fiable que los resultados que ofrece el modelo son parecidos a los que ofrece la realidad, para garantizar que el modelo representa de forma adecuada la realidad.
- Es necesario disponer de una metodología que permita estudiar configuraciones alternativas del sistema.
- En general, también puede ser interesante analizar los factores que condicionan el comportamiento de un sistema (diseño de experimentos, superficies de respuesta).

En este capítulo se ofrecen algunas nociones y técnicas básicas de estadística necesarias para la realización de un estudio de simulación de forma correcta.

2.2. Variables aleatorias

Variables aleatorias

Una de las características más notables de la simulación es la existencia de fenómenos no deterministas que se deben representar mediante variables aleatorias.

De una variable determinista se sabe con certeza el valor que toma. Por el contrario, de una variable aleatoria no se sabe con certeza el valor que toma, pero se conoce que puede tomar valores dentro de un determinado rango, de tal manera que existe una determinada probabilidad de que la variable tome un determinado valor dentro de dicho rango o se conoce la probabilidad de que dicha variable tome un valor determinado o uno menor que dicho valor.

De acuerdo con el tipo de valores que toma una determinada variable aleatoria, se pueden diferenciar entre:

- Continuas. Por ejemplo, la distribución normal, o la distribución exponencial.
- Discretas. Por ejemplo, la distribución de Poisson, o la distribución binomial.

Según el origen de los datos, se puede distinguir entre variables:

- Empíricas, en la que la probabilidad asignada a cada posible valor de la variable aleatoria se formula a partir de observaciones del propio sistema objeto de estudio.
- Teóricas, donde la probabilidad anterior se formula en términos analíticos y no procede de ningún conjunto de observaciones de un sistema real.

Función de distribución, función de probabilidad y función de densidad

Para una determinada variable aleatoria, se pueden ofrecer dos tipos de funciones para caracterizar el comportamiento de dicha variable aleatoria:

- Acumulada. Dada una variable aleatoria X , la función de distribución acumulada, conocida como *función de distribución*, relaciona cada posible valor de la variable aleatoria con la probabilidad de que dicha variable aleatoria tome un valor menor o igual que aquél. Es decir:

$$F(x) = p(X \leq x)$$

- Puntual. Según se trate de una variable discreta o continua, se habla de *función de probabilidad* o de *función de densidad*, respectivamente.

La función de distribución de una determinada variable aleatoria discreta X ofrece la probabilidad de que la variable tome un determinado valor, es decir:

$$f(x) = p(X = x)$$

Por su parte, dada una variable aleatoria continua X , se define la función de densidad $f(x)$ de la siguiente manera:

1. $f(x) \geq 0, \forall x$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $p(X \leq x) = \int_{-\infty}^x f(t)dt = 1$

Para el caso de las variables discretas, la relación entre la función de distribución y la función de probabilidad es la siguiente:

$$F(x) = \sum_{x_i < x} p(X = x_i)$$

En la figura 6 se muestra un ejemplo de las gráficas de las funciones de distribución y de probabilidad de una variable aleatoria discreta y la relación entre las mismas.

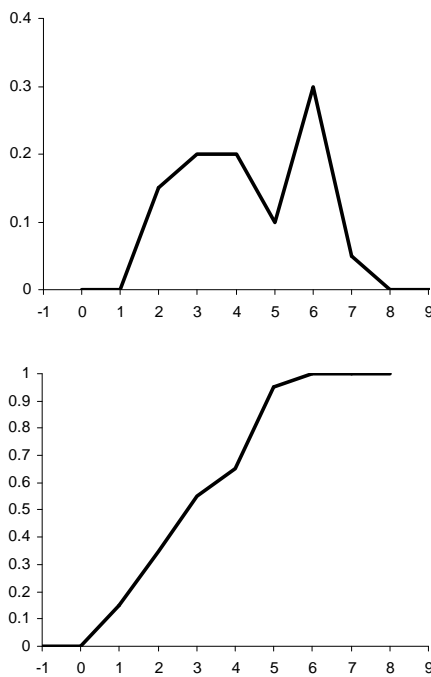


Fig. 6. Función de probabilidad y función de distribución de una variable aleatoria discreta

En el caso de las variables aleatorias continuas, la relación entre la función de distribución y la función de densidad es la siguiente:

$$F(x) = \int_{-\infty}^x f(t)dt = 1$$

En la figura 7 se muestra un ejemplo de las gráficas de las funciones de distribución y de probabilidad de una variable aleatoria continua y la relación entre las mismas.

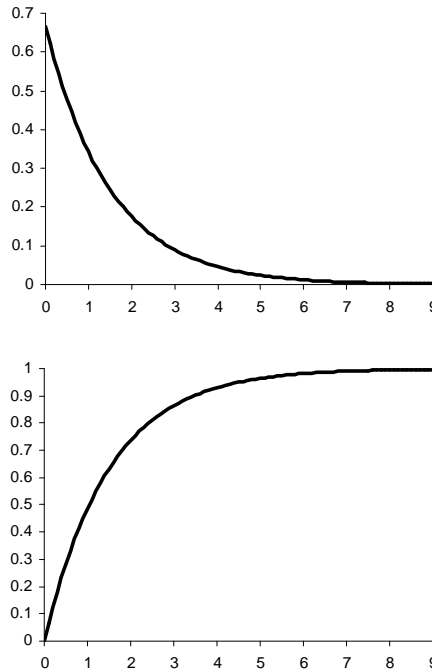


Fig. 7. Función de probabilidad y función de densidad de una variable aleatoria discreta

2.3. Media y varianza

Existen diferentes parámetros que resultan interesantes para caracterizar variables aleatorias. En particular, se presentan a continuación los dos más notables: la *media o esperanza matemática*, que permite caracterizar la tendencia central de la variable y la *varianza* que permite caracterizar la dispersión de los valores alrededor de la media. La definición es ligeramente distinta, según se trate de una variable discreta o una variable continua.

En particular, la esperanza matemática, $E(x)$ se define como:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \text{ si la variable es continua}$$

$$E(X) = \sum_i x_i f(x_i), \text{ si es discreta}$$

La varianza, $var(X)$, se define de la siguiente manera:

$$var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx, \text{ si la variable es continua}$$

$$var(X) = \sum_i (x_i - E(X))^2 f(x_i), \text{ si es discreta}$$

Poner otros parámetros que permitan caracterizar el las variables aleatorias.

2.4. Variables aleatorias más comúnmente utilizadas

A continuación, presentamos las distribuciones de probabilidad más comúnmente empleadas en la simulación. Para cada una de ellas, en primer lugar, se indican algunas de las posibles aplicaciones de las distribuciones, y se definen las funciones de densidad y de distribución. Después, se describen los parámetros de cada distribución, incluyendo sus posibles valores, y el rango al que las variables aleatorias asociadas pueden pertenecer y, finalmente, se incluyen las expresiones de la media (valor esperado) y de la varianza.

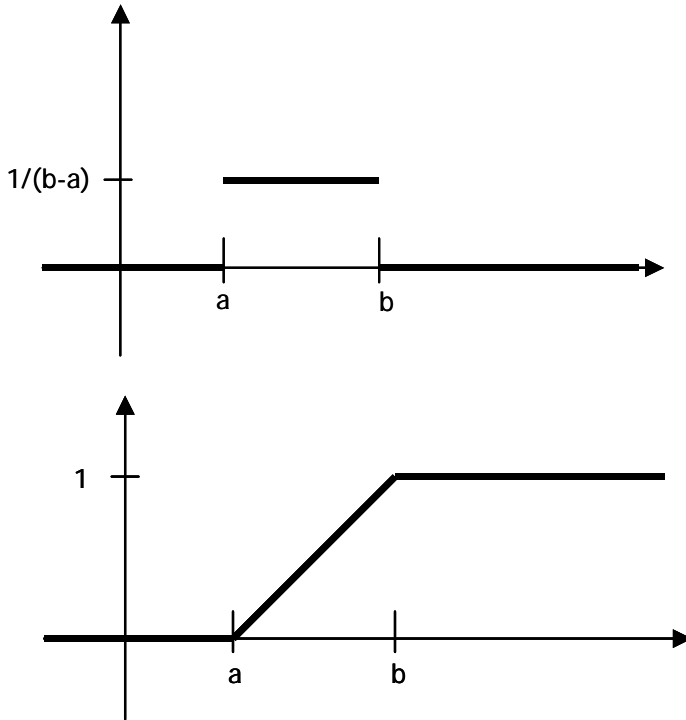
Distribuciones continuas más frecuentemente utilizadas

- Uniforme
- Exponencial
- Gamma
- Weibull
- Normal
- Normal-logarítmica
- Beta
- Triangular

Uniforme, $U(a,b)$

Esta distribución es generalmente u

Utilizada como una primera aproximación a una variable que varía aparentemente de forma uniforme entre dos valores, a y b . Como se verá más adelante, la variable $U(0,1)$ es la base para la generación de variables aleatorias.



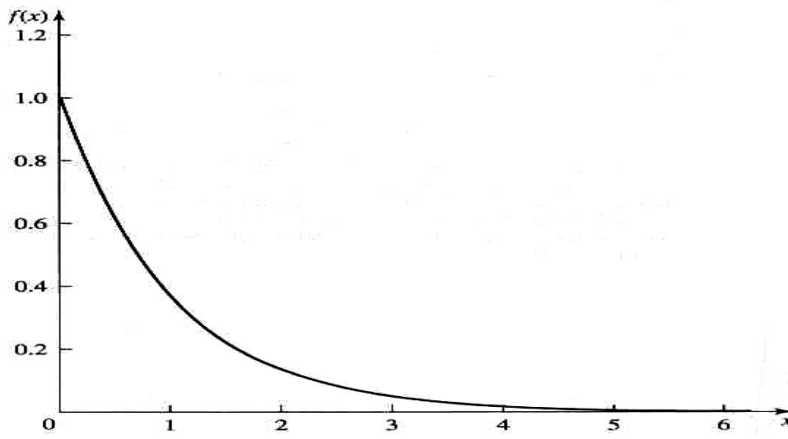
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

- Rango: $[a,b]$
- Media: $(a+b)/2$
- Varianza: $(b-a)^2/12$

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

Exponencial, Exp (β)

La variable exponencial representa, por ejemplo, el tiempo entre llegadas de clientes a un sistema que suceden a una tasa constante, o el tiempo transcurrido entre fallos de una máquina.



- Rango: $[0, \infty)$
- Media: β
- Varianza: β^2

Ej: $f(x)$ con $\beta=1$

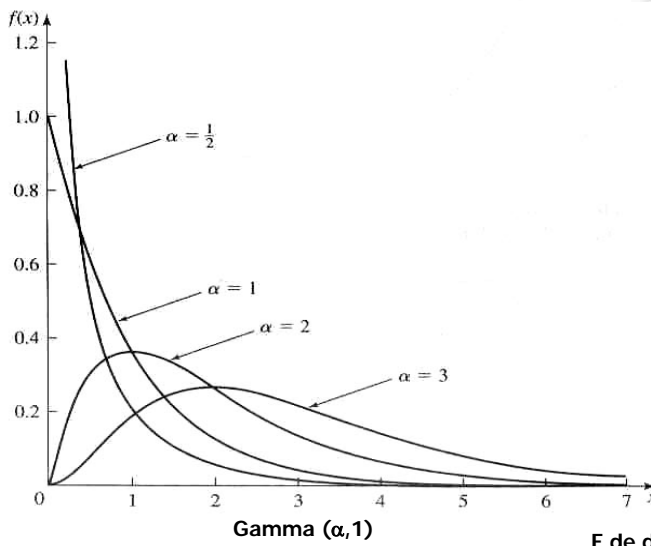
$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

1.

Gamma (α, β)

Esta variable puede permitir representar, por ejemplo, el tiempo para completar una tarea, como por ejemplo, el tiempo de servicio a clientes o de reparación de una máquina.



- Rango: $[0, \infty)$
- α, β positivos
- Media: $\alpha \beta$
- Varianza: $\alpha \beta^2$

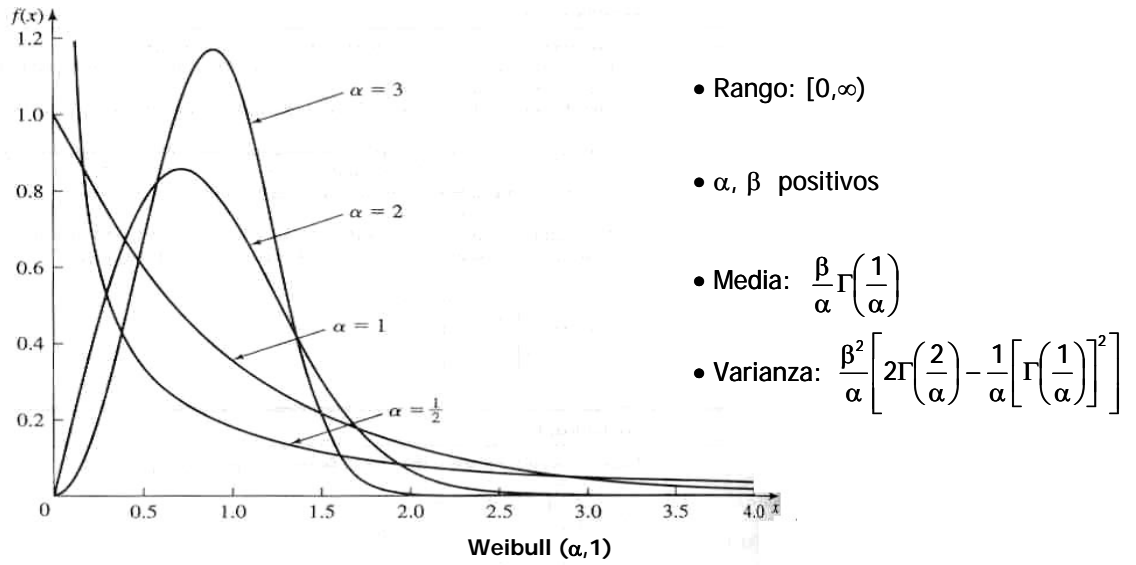
F de distribución: si $\alpha < 1$ no tiene forma cerrada, si α es un entero positivo:

$$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Weibull (α, β)

Puede servir para representa, por ejemplo, el tiempo para completar una tarea o el tiempo hasta el fallo de una máquina.

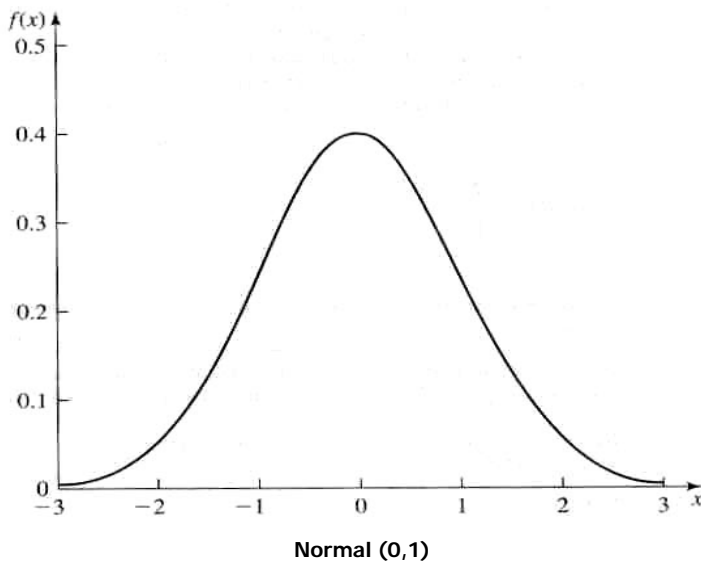


$$f(x) = \begin{cases} \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Normal (μ, σ^2)

Generalmente, representa errores de distintos tipos, o cantidades que son la suma de un gran número de otras cantidades.

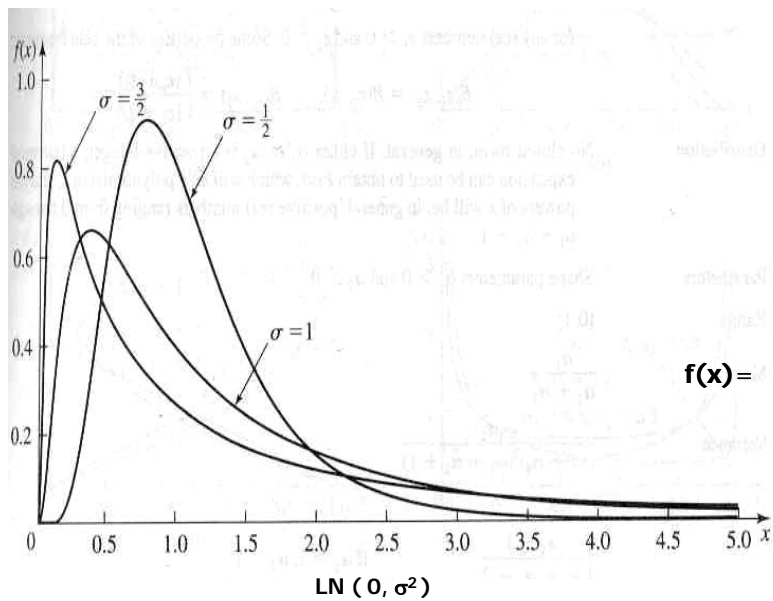


- Rango: $(-\infty, \infty)$
- μ no acotada y σ positiva
- Media: μ
- Varianza: σ^2

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

Normal-logarítmica (μ, σ^2)

Representa, entre otros, el tiempo para realizar una tarea, o cantidades que son el producto de un gran número de otras cantidades.

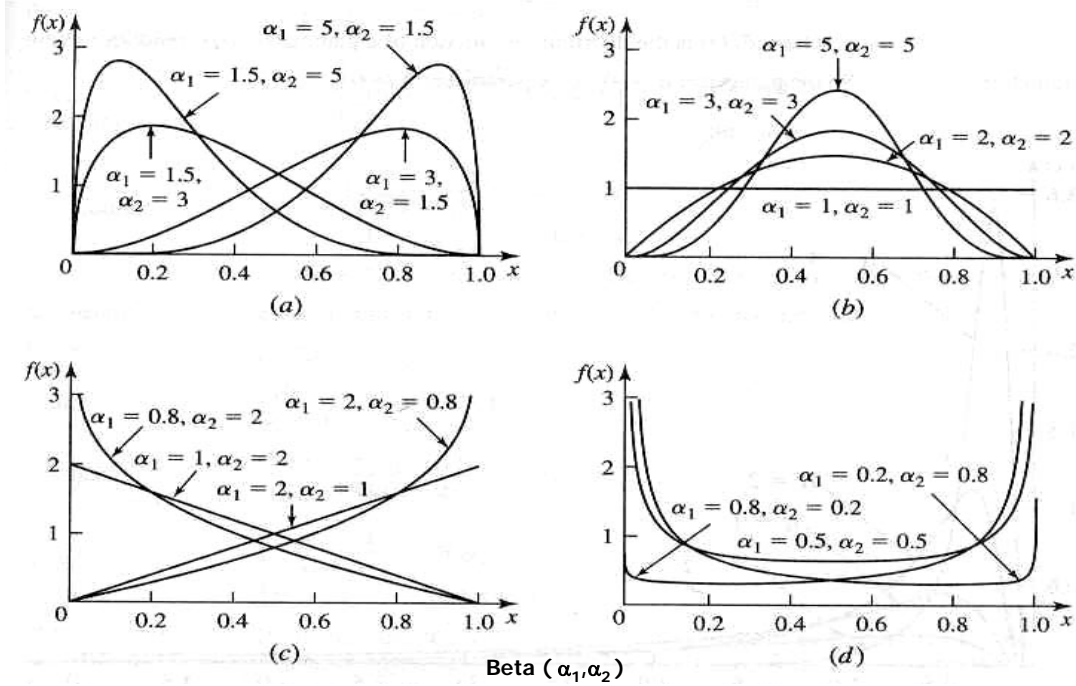


- Rango: $[0, \infty)$
- μ no acotada y σ positiva
- Media: $e^{\mu + \sigma^2/2}$
- Varianza: $e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1)$

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Beta (α_1, α_2)

Se utiliza para el modelado aproximado en ausencia de datos, o para representar la distribución del número de piezas de defectuosas en un lote, o el tiempo para completar una tarea.

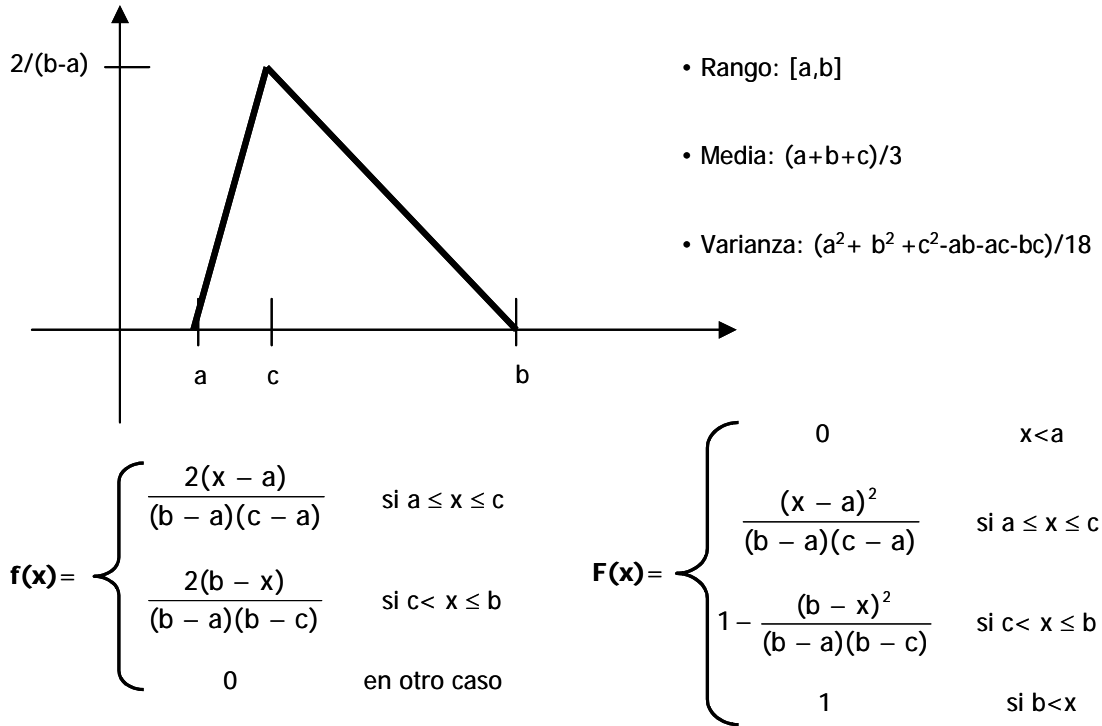


- α_1, α_2 positivos
- Rango: $[0, 1]$
- Media: $\frac{\alpha_1}{\alpha_1 + \alpha_2}$
- Varianza: $\frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}$

$$f(x) = \begin{cases} \frac{x^{\alpha_1-1} (1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Triangular (a, b, c)

Utilizada como una primera aproximación a una variable en ausencia de datos.



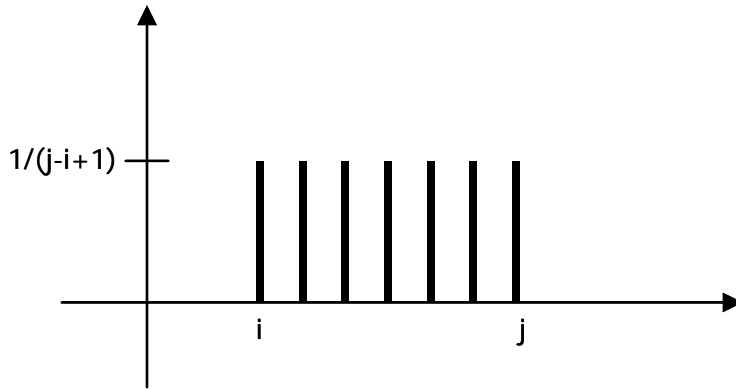
Distribuciones discretas más frecuentemente utilizadas

- Bernouilli
- Uniforme discreta
- Binomial
- Poisson

En este documento hablaremos sólo de las distribuciones uniforme discreta y de la Poisson.

Uniforme discreta (i, j)

Se emplea, por ejemplo, para el modelado aproximado en ausencia de datos que aparentemente varían entre dos valores extremos.



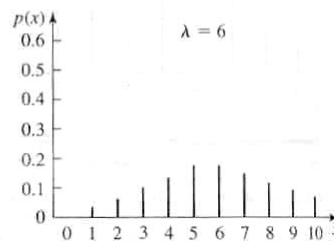
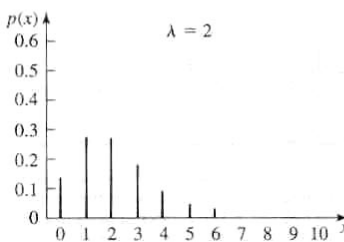
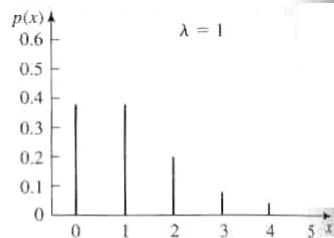
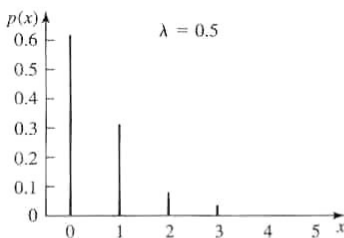
- Rango: $\{i, i+1, \dots, j\}$
- Media: $(i+j)/2$
- Varianza: $((j-i+1)^2-1)/12$

$$f(x) = \begin{cases} \frac{1}{j-i+1} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{si } x < i \\ \frac{|x| - i + 1}{j - i + 1} & \text{si } i \leq x \leq j \\ 1 & \text{si } j \leq x \end{cases}$$

Poisson (λ)

Representa el número de eventos que ocurren en un intervalo de tiempo cuando los eventos ocurren a una tasa constante, como por ejemplo el



- Rango: $\{0, 1, \dots\}$
- Media: λ
- Varianza: λ

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{si } x \geq 0 \end{cases}$$

número de piezas fabricadas a la hora.

2.5. Estimación de parámetros

Existen diversos métodos para la estimación de parámetros. A continuación se presenta el método del estimador máximo verosímil.

Si se supone que un conjunto de datos provienen de una determinada variable aleatoria continua, cuya función de densidad es $f_{\theta}(x)$, donde θ es el parámetro de la distribución, el estimador máximo verosímil de un conjunto de observaciones x_1, \dots, x_n es aquél que hace mínimo el valor de la función máximo verosímil, $\Gamma(\theta)$, que se define de la siguiente manera:

$$\Gamma(\theta) = f_{\theta}(x_1) \cdot f_{\theta}(x_2) \cdots f_{\theta}(x_n) = \prod_{i=1}^n f_{\theta}(x_i)$$

De forma análoga, si se trata de una variable aleatoria discreta cuya función de probabilidad es p_{θ} , la función toma la forma:

$$\Gamma(\theta) = p_{\theta}(x_1) \cdot p_{\theta}(x_2) \cdots p_{\theta}(x_n) = \prod_{i=1}^n p_{\theta}(x_i)$$

La función de verosimilitud representa la probabilidad de que (si la variable estudiada se comportara de acuerdo con la función de probabilidad p_{θ}) al generar n valores de dicha variables se obtengan exactamente los n valores x_1, \dots, x_n .

La forma de obtener el estimador máximo verosímil, por lo tanto, consiste en construir la función máximo verosímil, e igualar a cero su derivada con respecto al parámetro. Es decir, el estimador máximo verosímil, denotado por $\hat{\theta}$, verifica que:

$$\left. \frac{\partial \Gamma(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$$

Para el caso de una distribución exponencial, la función de verosimilitud, dado un conjunto de observaciones x_1, \dots, x_n , es la siguiente:

$$\Gamma(\beta) = f_{\beta}(x_1) \cdot f_{\beta}(x_2) \cdots f_{\beta}(x_n) = \frac{1}{\beta} e^{-\frac{x_1}{\beta}} \cdot \frac{1}{\beta} e^{-\frac{x_2}{\beta}} \cdots \frac{1}{\beta} e^{-\frac{x_n}{\beta}} = \prod_{i=1}^n \frac{1}{\beta} e^{-\frac{x_i}{\beta}} = \frac{1}{\beta^n} e^{-\frac{1}{\beta} \sum_{i=1}^n x_i}$$

El valor de β para el cual la función de verosimilitud es el mismo para el cual toma el máximo el logaritmo de dicha función, de manera que, por sencillez, se calcula de esta segunda manera:

$$\frac{\partial \ln(\Gamma(\beta))}{\partial \beta} = \frac{\partial \ln\left(\frac{1}{\beta^n} e^{-\frac{1}{\beta} \sum_{i=1}^n x_i}\right)}{\partial \beta} = \frac{\partial\left(-n \ln \beta + \ln\left(-\frac{1}{\beta} \sum_{i=1}^n x_i\right)\right)}{\partial \beta} = \frac{n}{\beta} - \frac{\sum_{i=1}^n x_i}{\beta^2} \Rightarrow$$

$$\frac{n}{\beta} - \frac{\sum_{i=1}^n x_i}{\beta^2} = 0 \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^n x_i}{n}$$

De esta manera, se pueden obtener los estimadores máximo verosímiles de las siguientes funciones (que se ofrecen en la tabla 1).

FUNCIONES	ESTIMADORES MÁXIMO VEROSÍMILES
Uniforme, U (a,b)	$\hat{a} = \min_{1 \leq i \leq n} X_i$ $\hat{b} = \max_{1 \leq i \leq n} X_i$
Exponencial, Exp (β)	$\hat{\beta} = \bar{X}(n)$
Weibull (α, β)	$\hat{\alpha} = \left\{ \frac{\left(\frac{6}{\pi^2} \right) \left[\sum_{i=1}^n (\ln X_i)^2 - \left(\frac{\sum_{i=1}^n \ln X_i}{n} \right)^2 \right]}{n-1} \right\}^{-1/2}$ $\hat{\beta} = \left(\frac{\sum_{i=1}^n X_i^{\hat{\alpha}}}{n} \right)^{1/\hat{\alpha}}$
Normal (μ, σ^2)	$\hat{\mu} = \bar{X}(n)$ $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}(n))^2}{n}}$
Normal-logarítmica (μ, σ^2)	$\hat{\mu} = \frac{\sum_{i=1}^n \ln X_i}{n}$ $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (\ln x_i - \bar{x}(n))^2}{n}}$
Uniforme discreta (i,j)	$\hat{i} = \min_{1 \leq k \leq n} X_k$ $\hat{j} = \max_{1 \leq k \leq n} X_k$
Poisson (λ)	$\hat{\lambda} = \bar{X}(n)$

Tabla 1. Estimadores máximo verosímiles de los parámetros de las funciones de algunas variables aleatorias

2.6. Ajuste de datos. Test de la χ^2

El test de la χ^2 permite posible comprobar si un conjunto de observaciones son una muestra independiente de una distribución.

Para ello, es necesario seguir los siguientes pasos:

- Comprobación de la independencia de las observaciones
- Formulación de la hipótesis de contraste
- Determinación del estadístico de contraste

Comprobación de la independencia de las observaciones

Dadas las observaciones, es necesario comprobar, en primer lugar, que se trata de un conjunto de observaciones independientes, para lo cual puede ser suficiente una comprobación de carácter gráfico.

Si se representan los pares de puntos (x_i, x_{i+1}) con, se puede apreciar gráficamente si existe algún tipo de relación entre una observación y la inmediatamente anterior o no. Por ejemplo, si se obtiene una gráfica como la de la izquierda de la figura 1, se obtiene una nube de puntos suficientemente dispersa como para garantizar la independencia de las observaciones. En el caso de la gráfica de la derecha, claramente, las observaciones están relacionadas linealmente.

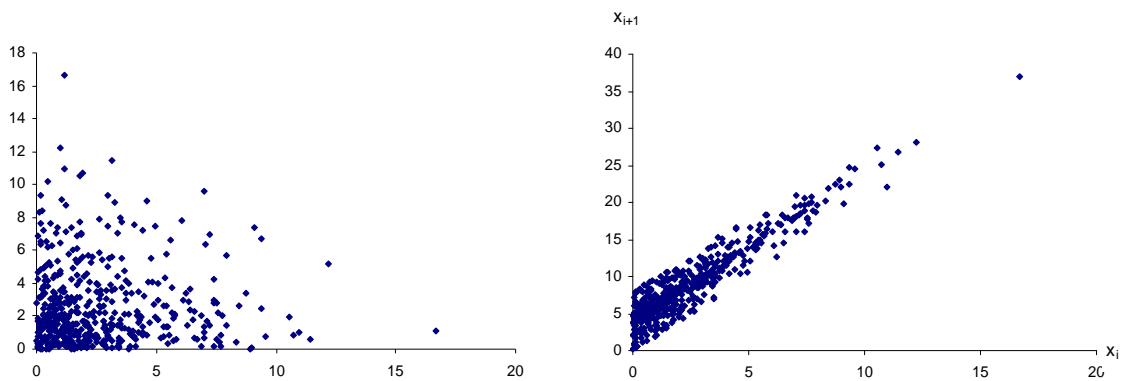


Fig. 1. Nubes de puntos correspondientes a observaciones no correlacionadas (izquierda) y a observaciones linealmente correlacionadas (derecha)

Formulación de la hipótesis de contraste

A partir de las observaciones es necesario realizar una hipótesis con respecto a la función de distribución que se puede ajustar de forma adecuada. Para formular dicha hipótesis, se realiza una representación gráfica de las observaciones. En concreto, se divide el rango de valores de las observaciones en un conjunto suficientemente grande de intervalos y se construye un histograma donde cada barra representa la frecuencia relativa de observaciones que pertenecen a cada intervalo.

A partir del histograma, se pueden formular hipótesis sobre la posible distribución. Por ejemplo, la figura 2 podría corresponder a una distribución exponencial, mientras que la figura 3 podría corresponder a una distribución Beta.

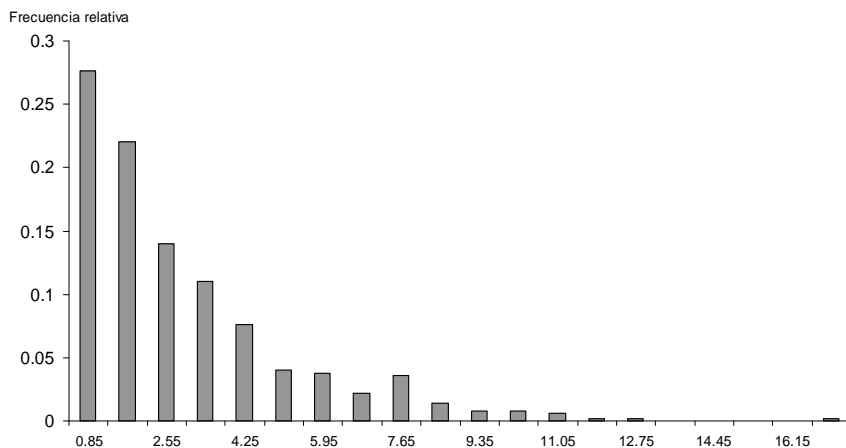


Fig. 2. Histograma de un conjunto de observaciones

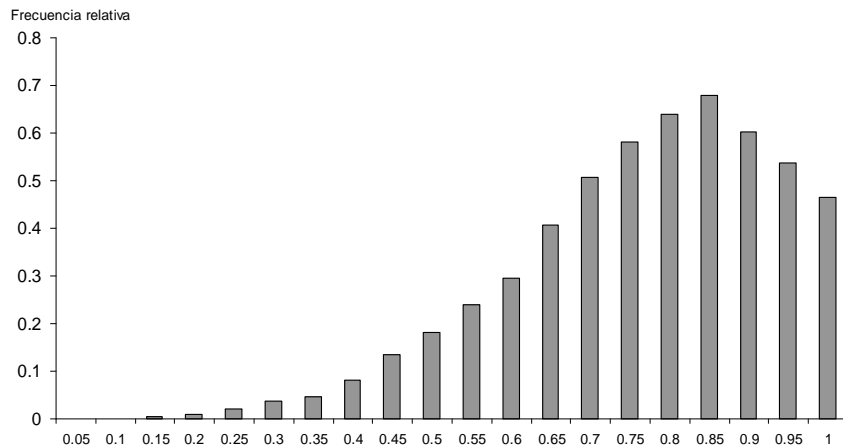


Fig. 3. Histograma de un conjunto de observaciones

Una vez seleccionada la función de distribución, es necesario caracterizarla, lo que significa estimar los parámetros de la misma. Por ejemplo, si a partir del histograma parece probable que se trata de una distribución exponencial, será necesario estimar el parámetro β , si se trata de una beta, será necesario estimar α_1 y α_2 .

En el apartado anterior se ha comentado cómo estimar los parámetros para una determinada variable a partir de un conjunto de datos.

Finalmente, se formula la hipótesis nula, H_0 , que se pretende contrastar:

Los datos x_1, \dots, x_n son valores independientes e idénticamente distribuidos correspondientes a una variable aleatoria X , con una función de densidad $f(x)$ de parámetros α, β, \dots

Determinación del estadístico de contraste

Una vez obtenido el parámetro de la función de distribución, se construye el estadístico de contraste. Para ello se divide el rango de valores de las observaciones en un conjunto de k intervalos de tal manera que la probabilidad de que una observación pertenezca a un determinado intervalo sea $1/k$, es decir, se trata de que los intervalos sean equiprobables. En general, la probabilidad de que una observación caiga en un determinado intervalo será p_i y, si es posible, se tratará de que $p_i = p_j = 1/k, \forall i, j$

Definimos la variable aleatoria E_i como el número de observaciones que caen dentro del i -ésimo intervalo cuando se generan n valores de una variable aleatoria X (la variable para la que se realiza el contraste). Dicha variable cumple las siguientes características:

- cada observación puede, o bien caer en el intervalo, o bien no caer;
- la probabilidad de que caiga dentro del intervalo es constante, p_i ; y
- la probabilidad de que una observación caiga o no en el intervalo es independiente de que caiga cualquier otra observación.

Por lo tanto, la variable aleatoria E_i , es una binomial de parámetros n, p_i :

$$E_i \sim B(n, p_i)$$

En caso de cumplirse que $np > 5$ y $n(1-p) > 5$, se puede aproximar la binomial por una normal:

$$E_i \sim N(np_i, \sqrt{np_i(1-p_i)})$$

Tipificando se obtiene que:

$$\frac{E_i - np_i}{\sqrt{np_i(1-p_i)}} \sim N(0,1) \sim Z$$

Elevando al cuadrado ambos lados se obtiene

$$\frac{(E_i - np_i)^2}{n(1-p_i)} \sim Z^2$$

Si $p_i \ll 1$, entonces se puede hacer la aproximación $np_i(1-p_i) \approx np_i$ y, por lo tanto:

$$\frac{(E_i - np_i)^2}{np_i} \sim Z^2$$

Finalmente, se sabe que la suma de k variables de tipo Z^2 es una variable aleatoria de tipo χ^2_{k-1} , donde k es el número de grados de libertad. Es decir:

$$\sum_j^k Z_j^2 \sim \chi^2_{k-1}$$

Si efectivamente, el conjunto de valores x_1, \dots, x_n , seguían una distribución de tipo X , el número de observaciones que caen en cada intervalo, denotadas por O_i , son valores en particular de las variables aleatorias E_i , de manera que el valor

$$\chi^2_{\text{exp}} = \sum_{j=1}^k \frac{(O_i - E_i)^2}{E_i}$$

es un valor de una variable χ^2_{k-1} . A este valor se le llama valor experimental.

Por otro lado, se puede calcular el valor $\chi^2(k-1, 1-\alpha)$, llamado estadístico teórico, que es el valor para el cual, la probabilidad de que una distribución χ^2 con $k-1$ grados de libertad tome dicho valor o uno menor sea $1-\alpha$. Con esta probabilidad, el valor experimental debe ser menor que el estadístico teórico.

Por lo tanto, si se cumple que $\chi_{\text{exp}}^2 > \chi^2(k-1, 1-\alpha)$, se rechaza la hipótesis nula, en caso contrario, no hay evidencia estadística para rechazarla.

2.7. Intervalos de confianza

Poner un ejemplo

Debido al carácter estocástico de las variables de entrada de los modelos de simulación, es natural, que las variables de salida sean, igualmente, variables aleatorias, de manera, que en diferentes ejecuciones del modelo se obtendrán diferentes valores para cada una de las variables. Un valor especialmente interesante es el de la media de dichas variables. Con el cálculo de un intervalo de confianza para la media de una determinada variable de salida, se obtiene un intervalo del que se puede afirmar que la media de la variable de salida está contenida en él con una determinada probabilidad. La información que ofrece dicho intervalo será tanto mayor cuanto menor sea su amplitud y cuanto mayor sea la probabilidad de que, efectivamente, contenga a la media.

Sea la variable aleatoria $Z_n = \frac{[\bar{X}(n) - \mu]}{\sqrt{\sigma^2/n}}$

y $F_n(z)$ su función de distribución para una muestra de tamaño n ; es decir:

$$F_n(z) = P(Z_n \leq z)$$

el teorema central del límite dice que, si n es “suficientemente grande”, entonces la variable aleatoria Z_n sigue una función de distribución Normal

[0,1], independientemente de la distribución que tenga la variable aleatoria X . Visto de otro modo, para una tamaño de muestra, n , elevado, la muestra de la media, $\bar{X}(n)$ sigue aproximadamente una distribución Normal, con media μ y varianza σ^2/n .

La dificultad de utilizar los resultados anteriores proviene de que, normalmente, la varianza σ es desconocida. Sin embargo, como la varianza de la muestra $S^2(n)$ converge hacia σ^2 cuando n aumenta, el teorema central del límite sigue siendo cierto si se sustituye σ^2 por $S^2(n)$ en la expresión de Z_n y, por lo tanto,

$$\frac{[\bar{X}(n) - \mu]}{\sqrt{\frac{S^2(n)}{n}}}$$

sigue una distribución $N[0,1]$.

Por lo tanto, acudiendo a las tablas de la distribución Normal se puede establecer un intervalo de confianza para el valor obtenido de la media. Es decir, si se establece un intervalo de confianza de $(1-\alpha)$ para μ , entonces:

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}(n) - \mu}{\sqrt{\frac{S^2(n)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = P\left(\bar{X}(n) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}} \leq \mu \leq \bar{X}(n) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}\right) \approx 1 - \alpha$$

Dicho de otro modo, existe una probabilidad de 100 $(1-\alpha)$ de que μ esté comprendido entre los valores:

$$\bar{X}(n) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}$$

Cuando el número n de observaciones no es “suficientemente elevado”, el teorema central del límite no se puede aplicar. En este caso, si se considera que las variables X_i siguen una distribución Normal, entonces la variable

$$t_n = \frac{[\bar{X}(n) - \mu]}{\sqrt{\frac{S^2(n)}{n}}}$$

Sigue una distribución t de Student con $n-1$ grados de libertad. En este caso, el procedimiento de establecer un intervalo de confianza para el valor de la media será similar al caso anterior, salvo que habrá que utilizar las tablas de la distribución t .

En la literatura sobre el tema se considera que un valor de $n \geq 30$ ya permite aplicar el teorema central del límite.

2.8. Comparación de alternativas

Ejemplos

Una vez visto el procedimiento para evaluar los resultados de un experimento de simulación, se está en disposición de comparar dos o más alternativas (es decir, los resultados de la simulación de dos o más sistemas alternativos) y seleccionar la mejor de ellas. Se pueden dar tres casos distintos:

- Comparación de dos alternativas.
- Comparación de varias alternativas con una de referencia.
- Comparación de varias alternativas y selección de la mejor de ellas.

Comparación de dos alternativas

Supongamos que se han realizado n repeticiones del experimento de simulación para cada una de las dos alternativas que se desea comparar.

Sean:

$$\begin{aligned} X_{11}, X_{12}, X_{13}, \dots, X_{1n} \\ X_{21}, X_{22}, X_{23}, \dots, X_{2n} \end{aligned}$$

los resultados obtenidos para las alternativas 1 y 2 respectivamente, y μ_1 y μ_2 las medias de sus variables correspondientes.

Se puede definir la variable Z como la diferencia entre los valores de la alternativa 1 y de la alternativa 2, es decir:

$$Z_j = X_{1j} - X_{2j} \quad \text{para} \quad j = 1, 2, 3, \dots, n$$

Estos valores de Z_j son variables aleatorias independientes y están idénticamente distribuidas.

A partir de la variable Z se puede construir un intervalo de confianza para la variable $\xi = \mu_1 - \mu_2$, es decir, para la diferencia de las medias de las variables estudiadas. Para ello, en primer lugar habrá que calcular la media y la varianza de Z :

$$\bar{Z}(n) = \frac{\sum_{j=1}^n Z_j}{n}$$

$$\text{Var}[\bar{Z}(n)] = \frac{\sum_{j=1}^n [Z_j - \bar{Z}(n)]^2}{n(n-1)}$$

Del mismo modo que se vio en el apartado anterior, se obtiene un intervalo de confianza de aproximadamente 100 $(1-\alpha)$ mediante la expresión:

$$\bar{Z}(n) \pm t_{n-1, 1-\frac{\alpha}{2}} \sqrt{\text{Var}[\bar{Z}(n)]}$$

Su interpretación es la siguiente: la comparación es significativa si el intervalo no contiene a 0 y, por el contrario, no es significativa si el intervalo contiene a 0.

Comparación de varias alternativas con una de referencia

Sea 1 la alternativa de referencia o base y 2, 3,...k el resto de alternativas que se quieren comparar con ella.

Es evidente que, utilizando reiteradamente el procedimiento descrito en el apartado anterior, se pueden realizar la comparación por separado de cada una de las alternativas 2,3,...k con la 1. Esto implica la comparación de k-1 parejas de alternativas.

Una opción al método anterior consiste en fijar simultáneamente intervalos de confianza, a un nivel global de (1- α) para las k-1 comparaciones. Para ello es necesario aplicar la desigualdad de Bonferroni, que dice que hay que construir intervalos individuales para las diferentes comparaciones de las medias ($\mu_2 - \mu_1, \mu_3 - \mu_1, \dots, \mu_k - \mu_1$) con un nivel $1 - \frac{\alpha}{k-1}$.

Una vez contruidos los k-1 intervalos de confianza individuales, si ningún intervalo de confianza para las diferencias de las medias ($\mu_i - \mu_1$) contiene el valor 0, entonces se puede decir que todas las alternativas difieren de la 1 con un nivel global de significancia de (1- α).

Si un intervalo de confianza para ($\mu_i - \mu_1$) contiene el valor de 0, se puede decir que no existe diferencia significativa de la alternativa i con respecto a la alternativa 1.

Selección de la mejor de k alternativas

Supongamos que se han realizado n repeticiones del experimento de simulación para cada una de las k alternativas que se desea comparar.

Sean

$$X_{11}, X_{12}, X_{13}, \dots, X_{1n}$$

$$X_{21}, X_{22}, X_{23}, \dots, X_{2n}$$

.....

$$X_{k1}, X_{k2}, X_{k3}, \dots, X_{kn}$$

los resultados obtenidos para las alternativas $1, 2, \dots, k$ respectivamente, y $\mu_1, \mu_2, \dots, \mu_k$ las medias de sus variables correspondientes, es decir, $\mu_i = E(X_{ij})$.

Si el objetivo de la selección es hallar la alternativa que proporcione un resultado menor, y se denomina μ_{i_1} al i -ésimo valor más pequeño de los μ_i , es decir:

$$\mu_{i_1} \leq \mu_{i_2} \leq \dots \leq \mu_{i_k}$$

el proceso de selección consistirá en hallar precisamente μ_{i_1} .

La aleatoriedad inherente a los valores X_{ij} obtenidos hace que no se pueda tener la absoluta seguridad de que la selección efectuada sea la correcta. Sin embargo, sí se puede especificar de antemano la probabilidad de que la selección que se haga sea la correcta.

Si los dos valores más pequeños obtenidos en las estimaciones de las medias de los experimentos, μ_{i_1} y μ_{i_2} , están muy cercanos, carecerá de importancia la selección errónea de μ_{i_2} en vez de μ_{i_1} . Por lo tanto, será conveniente utilizar un método de selección que evite hacer un número muy elevado de iteraciones para resolver una diferencia poco importante.

De acuerdo con las consideraciones anteriores, el objetivo del problema será seleccionar una alternativa que, con una probabilidad mínima P^* cumpla la condición $\mu_{i_2} - \mu_{i_1} \geq d^*$. El procedimiento que se indica a continuación tiene la propiedad de que, con una probabilidad de, al menos P^* , la respuesta esperada de la alternativa seleccionada no será mayor que $\mu_{i_1} + d^*$. Esto quiere decir que existe una protección (con una probabilidad de, al menos P^*) contra la selección de una alternativa cuya media sea una cantidad d^* peor que la del mejor de los sistemas.

El procedimiento precisa la especificación por parte del analista de los valores de P^* y d^* e implica un muestreo en dos etapas para cada uno de los k sistemas analizados.

Primera etapa

Consiste en realizar un número n_0 fijo de repeticiones o repeticiones para cada alternativa y utilizar los resultados obtenidos para estimar cuántas repeticiones más hay que efectuar en la segunda etapa. Es necesario asumir que los valores X_{ij} están normalmente distribuidos, pero no que los valores de $\sigma_i^2 = Var(X_{ij})$ sean conocidos, ni que los valores de σ_i^2 sean iguales para diferentes alternativas i .

Se hacen, por lo tanto, n_0 repeticiones de cada una de las k alternativas y se definen las medias y las varianzas de la primera etapa del siguiente modo:

$$\bar{X}_i^{(1)}(n_0) = \frac{\sum_{j=1}^{n_0} X_{ij}}{n_0} \quad \text{para } i=1,2,\dots,k$$

$$S_i^2(n_0) = \frac{\sum_{j=1}^{n_0} [X_{ij} - \bar{X}_i^{(1)}(n_0)]^2}{n_0 - 1}$$

A continuación, se calcula el número total de repeticiones, N_i , que es necesario realizar para cada alternativa i mediante la expresión:

$$N_i = \max \left\{ n_0 + 1, \left\lceil \frac{h_1^2 S_i^2(n_0)}{(d^*)^2} \right\rceil \right\}$$

donde:

$\lceil x \rceil$ indica el menor número entero que es mayor o igual al número real x .

h_1 se obtiene de la tabla siguiente, y depende de k , P^* y n_0 :

P^*	n_0	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
0.90	20	1.896	2.342	2.583	2.747	2.870	2.969	3.051	3.121	3.182
0.90	40	1.852	2.283	2.514	2.669	2.785	2.878	2.954	3.019	3.076
0.95	20	2.453	2.872	3.101	3.258	3.377	3.472	3.551	3.619	3.679
0.95	40	2.386	2.786	3.003	3.150	3.260	3.349	3.422	3.484	3.539

Tabla 2. Comparación de alternativas

Segunda etapa

Para cada alternativa i , se realizan $N_i - n_0$ repeticiones más y se hallan sus medias:

$$\overline{X}_i^{(2)}(N_i - n_0) = \frac{\sum_{j=n_0+1}^{N_i} X_{ij}}{N_i - n_0}$$

A continuación se definen los siguientes pesos:

$$W_{i1} = \frac{n_0}{N_i} \left[1 + \sqrt{1 - \frac{N_i}{n_0} \left(1 - \frac{(N_i - n_0)(d^*)^2}{h_i^2 S_i^2(n_0)} \right)} \right] \quad \text{para}$$

$$W_{i2} = 1 - W_{i1}$$

$$i=1,2,\dots,k$$

Por último, se definen las medias ponderadas de las muestras como:

$$\overline{X}_i(N_i) = W_{i1} \overline{X}_i^{(1)}(n_0) + W_{i2} \overline{X}_i^{(2)}(N_i - n_0)$$

y se selecciona la alternativa con un valor menor de $\overline{X}_i(N_i)$.

Es de destacar que, en la literatura sobre el tema, se recomienda que el número de repeticiones en la primera etapa, n_0 , sea como mínimo 20, ya que, si no es así, se obtiene una deficiente estimación de la varianza de la variable.

En el sentido opuesto, si n_0 toma un valor muy elevado, se pueden realizar más repeticiones de las necesarias para alguna alternativa, con el consiguiente coste de tiempo y recursos que ello implica.

3. Formulación del problema y modelo conceptual

Al realizar un estudio de simulación es necesario definir el sistema (es decir, el conjunto de elementos que se van a incluir en el modelo y, por exclusión, lo que queda fuera del modelo) y los objetivos que se pretenden conseguir con el estudio. Estas tareas son especialmente importantes por las siguientes razones:

- En primer lugar, para que el nivel de detalle del modelo sea el adecuado. No conviene que el detalle sea tan pobre que no se pongan de manifiesto los fenómenos relevantes del sistema, pero tampoco es interesante que el nivel de detalle sea mayor del necesario. En el primer caso, aunque el desarrollo pueda ser no muy costoso, los resultados serán poco fiables y el modelo inútil. En el segundo, el nivel de detalle puede no aportar información adicional interesante, implicará casi con toda seguridad un tiempo total de desarrollo mayor y, finalmente, se traducirá en un modelo informático más lento.
- En segundo lugar, la definición clara de los objetivos y la definición precisa del problema facilita la generación de alternativas potencialmente más interesantes y evita el estudio de alternativas menos atractivas.
- Finalmente, permite establecer las variables de salida adecuadas, para evaluar de forma correcta la eficacia y la eficiencia del sistema en diferentes circunstancias, además de facilitar que su estudio en términos estadísticos sea el adecuado.

En la elaboración del modelo conceptual (y para su posterior inclusión en el modelo informático), deben quedar claramente definidos los parámetros de diseño y las variables tanto de entrada como de salida.

Variables de entrada. Las variables de entrada son aquellas que corresponden a fenómenos del sistema real sobre las que no se tiene ningún

control. Por ejemplo, el número de clientes que llegan a una gasolinera, el producto que demandan, el volumen de demanda de un producto, etc.

Parámetros de diseño. Los parámetros de diseño representan aquellas características del sistema sobre las que se tiene control y que determinan la configuración del sistema. Pueden ser parámetros de diseño, el número de surtidores que se dedica a cada producto en una gasolinera, el número de operarios que se destina a cierto conjunto de operaciones, la máquina que se emplea para realizar una determinada operación.

Dependiendo de la naturaleza del estudio, un determinado fenómeno puede ser una variable de entrada o un parámetro de diseño. Por ejemplo, durante la fase de diseño de una línea de montaje, el tiempo de operación de una determinada máquina puede ser un parámetro de diseño, ya que en el estudio se consideran diferentes tipos de máquina, cada una de ellas con un tiempo de operación distinta. Si, en cambio, el estudio de la simulación se refiere a la definición de puestos de trabajo y asignación de responsabilidades a operarios en una línea ya montada y sin posibilidad de operación a corto plazo, el tiempo de operación de las máquinas se considerará que es una variable de entrada del modelo.

Variables de salida. Finalmente, a partir de las variables de entrada y de los parámetros de diseño del modelo y de las relaciones que se establecen entre diferentes elementos del modelo, se obtienen valores para las variables de salida, que son aquellas que, generalmente, permiten caracterizar el sistema para una determinada configuración y que están estrechamente ligadas a los objetivos del estudio. Por ejemplo, el tiempo total de espera en una oficina de atención puede ser una variable de salida. Igualmente, la producción diaria de una planta puede ser una variable de salida.

3.1. Recogida, análisis y generación de datos de entrada

Para simular el comportamiento de un determinado sistema se necesita alimentar el modelo con diferentes valores de las variables de entrada (por

ejemplo, los tiempos de llegada entre clientes a un banco, el número de piezas de cada pedido que llega a una planta de producción, etc.)

En ocasiones, no se dispone de datos históricos de las variables de entrada (porque no se han recopilado, porque no ha existido la oportunidad de recogerlos, etc.) Cuando esto ocurre, conviene alimentar el modelo con alguna distribución teórica de la que es razonable suponer que tiene algún parecido con la realidad. Por supuesto, los resultados serán tanto más fiables cuanto más acertada sea la elección.

Si, en cambio, se dispone de datos históricos, existen tres alternativas:

- Alimentar el modelo con los datos históricos tal y como se han recogido. Esta alternativa es interesante desde el punto de vista de la validación del modelo, es decir, para confirmar que el modelo representa de forma adecuada el sistema estudiado. Efectivamente, si se dispone de un conjunto de valores para las variables de entrada del modelo y de los correspondientes valores de salida que ofreció el sistema real para dichos valores, es posible comprobar si para dichos valores, el modelo arroja valores parecidos a los reales para las variables de salida.

Desde el punto de vista de la explotación del modelo de simulación, esta alternativa es muy poco interesante; como sólo se dispone de un conjunto finito de valores históricos, sólo es posible simular el comportamiento frente a dichos valores, es decir, sólo es posible reproducir lo que históricamente ha ocurrido. Además, tiene muy poco interés simular un sistema alimentándolo únicamente con valores históricos dado que se obtendrá, en el mejor de los casos, un conjunto de valores de salida parecidos a los que ofreció el sistema real (de los que ya se dispone.) y, en el peor de los casos, se obtendrán valores alejados de los valores reales.

- Construir una función de distribución empírica y generar valores de acuerdo con dicha distribución.

Esta alternativa consiste en dividir el rango de valores en un conjunto de intervalos o de valores discretos y asignar una probabilidad a cada

intervalo o cada valor proporcional a la frecuencia relativa de los valores históricos.

A diferencia de la anterior, esta alternativa permite obtener un número infinito de valores comprendidos entre el máximo y el mínimo de los valores históricos.

- Realizar un ajuste de los datos históricos a una función de distribución teórica, y generar valores de acuerdo con dicha distribución.

Desde el punto de vista de la calidad de la simulación, esta alternativa es la más interesante de las tres. Como en la anterior, es posible obtener un número infinito de valores. Sin embargo, con esta alternativa es posible obtener valores fuera del rango de los datos originales, con lo que se gana generalidad en el modelo.

Por otro lado, las posibles irregularidades de los datos históricos, se evitan.

Por último, desde el punto de vista práctico, las distribuciones teóricas son compactas y, por lo tanto, más sencillas de modificar, por lo que permiten introducir modificaciones en el modelo de simulación con más facilidad.

Generación de números aleatorios

Los números aleatorios, entendiendo como tales a los distintos valores de la variable aleatoria uniformemente distribuida en el intervalo $[0,1]$ son la base de partida para la generación de valores de cualquier variable aleatoria que siga una cierta función de distribución. Existen distintos procedimientos de generación de números aleatorios. El método que se presenta a continuación no proporciona una sucesión de números totalmente aleatorios, ya que la secuencia está perfectamente determinada a partir de los valores iniciales de los parámetros. No obstante, si se seleccionan adecuadamente estos valores, sus características a todos los efectos son prácticamente idénticas a las de los números aleatorios. Por este motivo se denominan

números *pseudo-aleatorios*. Presentan como ventaja el hecho de que no es necesario utilizar una gran cantidad de almacenamiento en el ordenador, ya que se van generando a medida que se van necesitando mediante un procedimiento recurrente.

Para la generación de una sucesión de números *pseudo-aleatorios* (en adelante los denominaremos aleatorios, ya que tienen su mismo comportamiento) en el intervalo $[0,1]$ se puede utilizar el procedimiento de Lehmer o de las congruencias. Este método consiste en un cálculo recurrente, en el cual un nuevo número aleatorio se obtiene del último generado aplicando la siguiente expresión:

$$Z_i = (a \cdot Z_{i-1} + b) \bmod(m), \text{ y } r_i = \frac{Z_i}{m}$$

donde:

r_i es el i -ésimo número aleatorio de la sucesión generada,

Z_i es el i -ésimo número natural obtenido en la secuencia

a , b y m son parámetros,

\bmod representa la operación módulo, que devuelve el resto de la división de los dos operandos entre los que aparece.

Conviene observar que, tal y como se generan números aleatorios con este método:

- Los posibles valores de r_i son: $0/m, 1/m, \dots, m-1/m$
- El número de valores diferentes que se pueden generar es finito y es, como máximo, igual al valor del parámetro m .
- El la medida en la que el valor de m sea mayor será más probable que el conjunto de valores diferentes sea mayor y, además, que la densidad del conjunto obtenido sea igualmente mayor.

Todos los lenguajes de ordenador de alto nivel llevan incorporada una rutina de generación de números aleatorios que utiliza este procedimiento. Tanto el valor de los parámetros como el valor del primer número de la sucesión están elegidos de tal manera que las sucesiones de números aleatorios que se obtienen garantizan suficientemente que los resultados obtenidos en la simulación no van a estar sesgados por la aparición de ciclos periódicos o por su concentración alrededor de ciertos valores.

Los lenguajes de propósito general suelen incluir generadores de bastantes variables aleatorias, aunque generalmente, no suelen ser generadores de gran calidad. Los paquetes de software comerciales incluyen un gran número de variables aleatorias. En caso de que la calidad de los generadores no sea suficientemente buena, existen rutinas que permiten programar la generación de números aleatorios de buena calidad. En Law y Kelton (1991) se pueden encontrar otros generadores congruentes lineales más sofisticados, así como otros generadores de naturaleza distinta. En Fisher (1978, 358-7171) y Knuth (1981, 75-110) se pueden encontrar tests que permiten evaluar la calidad de diferentes generadores de números aleatorios.

Generación de valores de variables aleatorias

Como ya se ha indicado, la generación de una sucesión de valores de una variable aleatoria se realizará partiendo de una sucesión de números aleatorios generados previamente. A continuación se indica la generación de estos valores para varias funciones de distribución utilizadas frecuentemente en simulación. En dichas expresiones, r indica siempre un número aleatorio, y x el valor de la variable aleatoria.

Función de distribución uniforme en el intervalo [a,b]

$$x = a + (b - a)r$$

Función de distribución exponencial

$$x = -\beta \ln r$$

donde β es la media de la distribución.

Función de distribución normal

Uno de los métodos más usados es el denominado polar, que consiste en:

1. Tomar dos números aleatorios r_1 y r_2 , y hacer:

$$v_1 = 2r_1 - 1$$

$$v_2 = 2r_2 - 1$$

Calcular: $w = v_1^2 + v_2^2$

2. Si $w > 1$, volver al paso 1, si no, hacer:

$$y = \sqrt{(-2 \ln w) / w}$$

$$x_1 = v_1 y$$

$$x_2 = v_2 y$$

Entonces, x_1 y x_2 siguen una función de distribución Normal (0,1).

Para generar valores de una función de distribución Normal x' con otra media y otra desviación típica, $N(\mu, \sigma^2)$, bastará con hacer el siguiente cambio de variable:

$$x' = \mu + \sigma x$$

Función de distribución lognormal

1. Hacer:

$$\mu = \ln(\mu_i^2 / \sqrt{\sigma_i^2 + \mu_i^2})$$

$$\sigma^2 = \ln[(\sigma_i^2 + \mu_i^2) / \mu_i^2]$$

2. Generar un valor aleatorio y de la variable $N(\mu, \sigma^2)$
3. $x = e^y$

Función de distribución de Weibull

$$x = \beta(-\ln r)^{1/\alpha}$$

siendo α y β los parámetros de la distribución.

Función de distribución discreta

Consideremos la situación general en la cual conocemos las probabilidades $p(0), p(1), p(2), \dots$ de S números enteros no negativos, y queremos generar sucesivos valores de una variable aleatoria discreta x que siga la correspondiente distribución. Los distintos valores $p(i)$ pueden haberse establecido teóricamente o bien empíricamente mediante una observación directa de un determinado fenómeno. Los pasos son los siguientes:

1. Generar r
2. Hacer $x=I$, de tal forma que se satisfaga la siguiente expresión:

$$\sum_{j=0}^{I-1} p(j) \leq r < \sum_{j=0}^I p(j)$$

Los paquetes comerciales de simulación permiten gobernar la generación de números aleatorios para garantizar que los valores de variables aleatorias correspondientes a fenómenos independientes sean, efectivamente independientes. Por ejemplo, en un modelo de simulación de

una gasolinera, el tiempo entre llegada de vehículos sigue una exponencial de tres minutos de media y el tiempo de llenado del depósito sigue una normal de media tres minutos y varianza 1.5. Los números aleatorios a partir de los cuáles se generan valores de dichas variables, deben ser independientes.

En particular, por ejemplo, Witness permite la introducción de lo que se denomina “semilla”. En la figura X se muestra el cuadro de diálogo en el que se introduce el tiempo entre llegada de clientes al sistema; en la llamada a la función exponencial que aparecen en la caja de texto titulada ‘Inter Arrival Time’ aparecen dos parámetros: uno de ellos es la media (3) y el otro es la semilla (7). Una vez fijada la semilla, se fija la secuencia de números aleatorios que se utilizan para la generación de valores. La semilla que se debe utilizar en la generación de tiempos de servicio debe ser distinta, para evitar que se el tiempo de servicio y el tiempo entre llegada de vehículos esté ligada. Igualmente, es necesario modificar el valor de la semilla correspondiente a diferentes replicaciones, o de lo contrario, los valores obtenidos en dichas replicaciones serían exactamente los mismos.

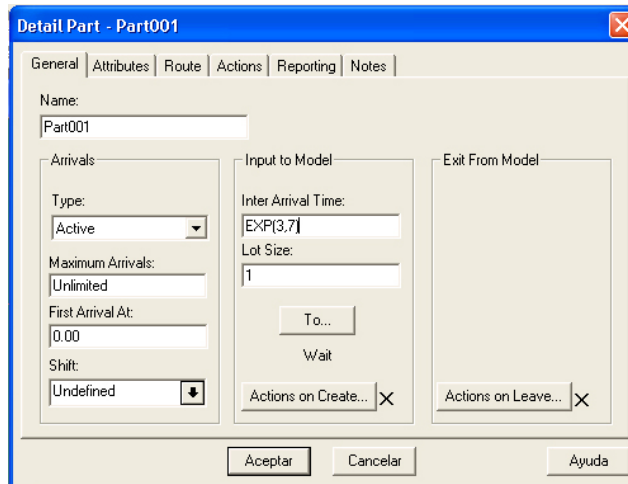


Fig X. Generación de variables aleatorias en Witness: semillas

4. Construcción del modelo. Verificación, validación y credibilidad

Existen tres características que un modelo de simulación debe ofrecer para servir a su propósito; debe ser válido, veraz y creíble.

Un modelo es tanto más válido cuanto mejor representa el sistema objeto de estudio con respecto a los objetivos del estudio. Por un lado, se valida el modelo conceptual, previamente al desarrollo del modelo informático. Por otro, una vez elaborado este, se debe comprobar que, efectivamente, el modelo informático representa de forma adecuada el sistema.

Para que un modelo sea válido, debe establecerse un nivel de detalle adecuado, conviene explicitar el modelo conceptual, se debe recoger información relevante y precisa, puede ser interesante recopilar información de los gestores, se deben analizar las hipótesis tanto implícitas como explícitas. Además de lo anterior, si se dispone de datos históricos correspondientes tanto a las variables de entrada como a las de salida, se puede ejecutar el modelo con los datos de las variables de entrada y comprobar si los valores de las variables de salida son parecidos a los valores que se obtuvieron en realidad.

La verificación de un modelo consiste en la realización de actividades orientadas a garantizar la correcta programación del modelo de simulación. La verificación está íntimamente ligada con el entorno de simulación elegido, para la cual existen herramientas muy útiles. Para realizar una correcta verificación puede ser conveniente, por ejemplo, utilizar un enfoque modular para estudiar el comportamiento de cada módulo por separado. Igualmente, la tarea de verificación es más sencilla si aumenta de forma progresiva la complejidad del modelo, verificando previamente cada modelo antes de introducir más elementos de complejidad. También puede ser interesante ejecutar el modelo bajo hipótesis simplificadas.

La credibilidad de un modelo radica en la confianza que los gestores depositan él para tomar decisiones relativas al sistema. Por supuesto, la validez y la veracidad de un modelo contribuyen favorablemente a su credibilidad, pero no la garantizan. Un estudio de simulación en el que no ha existido colaboración por parte de los gestores, o estos se han mostrado desconfiados, por ejemplo, es muy probable que reciba poca atención. Conviene prestar atención a lo largo del proceso a este aspecto, para evitar que el estudio sea estéril.

4.1. Entornos de simulación.

Herramientas informáticas disponibles

Una vez que se ha construido y validado el modelo de simulación, se debe seleccionar el lenguaje que se va a utilizar para su programación.

El software disponible para el desarrollo de modelos de simulación puede ser dividido en cuatro categorías.

- **Hojas de cálculo.** Cuando se trabaja con problemas de pequeña dimensión es posible usar también hojas de cálculo, como por ejemplo Excel, para tener una idea del funcionamiento de un sistema. Para problemas de relativa complejidad, las hojas de cálculo se vuelven demasiado engorrosas y dejan de ser una alternativa interesante.
- **Lenguajes de propósito general.** Fueron muy empleados en el nacimiento de la simulación pero requieren mucho tiempo de programación y, por eso, se prefiere, en general, usar lenguajes específicos para la simulación. A esta categoría pertenecen lenguajes como Fortran, C y C++.
- **Lenguajes para la programación de simulación.** Proporcionan muchas características necesarias para realizar un modelo de simulación, reduciendo así el tiempo de realización. Ejemplos son GPSS, Simscript, Siman, Modsim, etc. Aunque son menos flexibles que los

lenguajes de propósito general, son el modo más natural para realizar un modelo de simulación.

- **Paquetes de software de simulación.** Los entornos específicos de simulación son herramientas muy potentes con elementos predefinidos que simplifican notablemente la tarea de construcción de modelos. Estos elementos permiten representar de forma sencilla un abanico relativamente amplio de procesos y fenómenos que tienen lugar en entornos logísticos y productivos. A pesar de ello, a veces es necesario diseñar elementos específicos o configurar de manera adecuada los elementos existentes para representar particularidades del sistema estudiado.

Dentro de este grupo, existen lenguajes como Extend, Micro Saint, Autmod, Promodel, Arena Y Witness.

Estos paquetes incluyen, además, herramientas específicas para gestionar todo el proceso de simulación (como por ejemplo, la realización de experimentos). El inconveniente de esta alternativa es el precio, notablemente mayor que el de cualquiera de las alternativas anteriores.

Witness©

[Comentar con detalle aquí Witness](#)

Arena©

[Comentar con detalle aquí Witness](#)

4.2. Verificación

La tarea de verificación consiste en el conjunto de acciones destinadas a garantizar una correcta programación del modelo, es decir, a conseguir que el modelo realice las acciones conforme se desea que tengan lugar de acuerdo con el modelo conceptual.

Como la verificación no sino la tarea de depuración de un programa informático con las características particulares de la simulación, las orientaciones generales propias de la depuración de un programa son válidas. Además, es recomendable tener en cuenta las siguientes recomendaciones.

Enfoque modular

En primer lugar, al construir el modelo a base de módulos se facilita la tarea de verificación, ya que es más sencillo comprobar la corrección de fragmentos de código no demasiado extensos.

Además, si los módulos representan subsistemas del sistema real, es posible verificar el funcionamiento de cada uno de ellos de manera independiente y, después, el conjunto del modelo una vez integrados los diferentes módulos.

De lo sencillo a lo complicado

Esperar a verificar el modelo hasta un momento en el que se haya avanzado en gran medida en el desarrollo del modelo es una mala práctica. Conviene, comenzar por realizar modelos sencillos, verificarlos y, progresivamente, aumentar su complejidad.

Por ejemplo, en la representación de una planta de producción, puede ser interesante representar diferentes secciones en la planta como módulos sin ningún tipo de detalle. Verificar el funcionamiento y, posteriormente, detallar cada una de las secciones. Igualmente, dentro de cada sección, se puede aumentar la complejidad (el nivel de detalle) de forma progresiva.

Contrastes de grupo

Es frecuente que los miembros de un equipo de programación se reúnan para revisar el código realizado por ellos. Con esto es posible detectar errores que algún programador pasa sistemáticamente por alto, o permite disponer de más alternativas de programación para elegir después la más eficiente o la más adecuada para su posterior ampliación o modificación.

Ejecución de una amplia variedad de configuraciones

Otra técnica de verificación consiste en evaluar un conjunto amplio de configuraciones del modelo. En particular, puede ocurrir que determinadas partes del modelo, para determinadas configuraciones, permita ser representado mediante un modelo exacto, de manera que es posible comparar los resultados que se obtienen con el modelo de simulación para dichas configuraciones y el resultado teórico del modelo exacto.

Por otro lado, conviene también estudiar el comportamiento del sistema en diferentes configuraciones (aunque no se disponga de modelos exactos para comparar). Por ejemplo, si se aumenta notablemente el tiempo de ciclo de una determinada máquina, cabe esperar que se acumule stock aguas arriba de la máquina y no aguas abajo. Se deben realizar comprobaciones sencillas de este tipo, que permitan confirmar el correcto funcionamiento del sistema.

Ejecución paso a paso

Todos los lenguajes de simulación incorporan herramientas para depurar los modelos realizados, unos más potentes que otros. Una estrategia para verificar el modelo consiste en ejecutarlo paso a paso y comprobar, tras cada modificación elemental del modelo, que el estado del sistema es que estaba previsto: las variables almacenan los valores correctos, los diferentes elementos que representan máquinas, almacenes, colas etc. están contienen las piezas adecuadas y en las cantidades previstas, etc.

Igualmente, los depuradores incorporan la posibilidad de ejecutar sentencias para modificar instantáneamente el valor de alguna variable o para cambiar el estado del modelo, en caso de que se desee que la situación del modelo sea una determinada para, posteriormente, comprobar si el modelo evoluciona correctamente a partir de dicha situación.

Ejecución con hipótesis simplificadas

Para identificar posibles errores en determinadas partes del modelo, se pueden simplificar las hipótesis de funcionamiento en elementos distintos de aquellos que se quieren verificar. Por ejemplo, para comprobar que la ruta que siguen las piezas a lo largo de un taller, puede ser más sencillo suponer que la capacidad de almacenamiento de producto intermedio es infinita sin que esto afecte a la lógica de la asignación de rutas. Una vez comprobado que las rutas se asignan correctamente, se puede restituir la hipótesis de almacenamiento limitado y continuar con la verificación de otros aspectos del modelo.

Herramientas del entorno de simulación

Conviene explotar convenientemente las herramientas que ofrece el propio entorno de simulación para programar de manera correcta el modelo. Por ejemplo, si las piezas procesadas por una determinada máquina se entregan en una determinada proporción a tres almacenes distintos, se puede programar una instrucción compleja mediante la generación de un número aleatorio y en función del valor de dicho número entregarlo al almacén correspondiente. Sin embargo, es muy probable que exista una sentencia mucho más compacta proporcionada por el propio entorno de simulación para ejecutar dicha acción. Con esta sentencia, primero, en caso de que se conozca su lógica correctamente, existen menos posibilidades de que se introduzcan errores, segundo, su modificación (dado que suelen ser más compactas) es más sencilla y, tercero, el tiempo de ejecución de este tipo de sentencias es típicamente menor que otras más complejas creadas por el programador.

Utilización de las herramientas de animación

Un caso particular de la explotación de las herramientas que ofrecen los entornos de simulación es la utilización de los elementos gráficos que permiten representar dinámicamente los diferentes estados por los que atraviesa el sistema, el movimiento de las entidades por los diferentes elementos, etc.

Algunos entornos ofrecen más facilidades para disponer de una animación fiable sin una inversión de tiempo muy elevada.

4.3. Validación

El proceso de validación tiene por objeto garantizar que el modelo de simulación desarrollado representa suficientemente bien el sistema real objeto del estudio. Existen técnicas estadísticas que permiten comparar los resultados que ofrece el modelo con los datos históricos del sistema real.

Además de las técnicas estadísticas, a continuación se presentan algunas recomendación que conviene tener presentes durante la realización del estudio de simulación para

Determinación del nivel de detalle adecuado y acorde con los objetivos del estudio.

Un modelo de simulación es una representación más o menos fiel de un sistema real, que se realiza con un objetivo determinado; se recurre al modelo como sustituto del sistema real por las razones que se han comentado en **X**.

Para que el modelo sea válido debe permitir simular la toma de decisiones que se pueden (o se podrán) tomar, de hecho, en el sistema real.

Por ejemplo, en el estudio de diferentes políticas para la inspección de las maletas a diferentes niveles en un aeropuerto, el detalle del modelo, debe permitir, por ejemplo, modificar la asignación de personal a los diferentes equipos de inspección, o se debe representar con suficiente detalle las cintas transportadoras para garantizar que las inspecciones se pueden realizar con los equipos disponibles. En casa contrario, no se podría evaluar el interés de unas cintas frente a otras o de unos equipos de inspección frente a otros. Por el contrario, si el estudio pretende estudiar la asignación de vuelos a las diferentes compañías, quizá no tenga tanto interés llegar al nivel de detalle anterior,

porque para decidir la política de asignación no es necesario incluir los equipos de inspección de equipajes.

Igualmente, en un sistema de tipo job-shop, si se desarrolla un modelo de simulación para estudiar diferentes alternativas de gestión de la producción para determinar el área necesaria para alojar los diferentes stocks de productos intermedios, es necesario explicitar el área que estos ocupan. Sin embargo, si de lo que se trata es de comparar diferentes reglas de programación, y las áreas destinadas a stock de productos intermedios no representan un problema, no es necesario explicitarlas, ya que en este caso no aportaría información relevante.

Es decir, a partir de cierto nivel de detalle, los resultados que se obtienen no son suficientemente mejores como para que el modelo resultante sea interesante, ya que, por una lado, un nivel de detalle excesivamente elevado implica un tiempo total de desarrollo del modelo muy elevado (y por lo tanto un coste igualmente elevado) y, en segundo, el tiempo de computación necesario para realizar una replicación puede ser innecesariamente alto. En definitiva, un nivel de detalle superior al necesario entraña mayores costes y mayores tiempos de desarrollo y ejecución.

Explicitación del modelo conceptual

Conviene mantener registradas todas las hipótesis tanto explícitas como implícitas asumidas durante el desarrollo del modelo.

Desde el punto de vista del periodo de desarrollo del modelo, conviene disponer de la documentación adecuada que permita a los implicados en el proceso conocer las hipótesis para evaluar, modificar, etc. el modelo de manera consistente. Si el enfoque del desarrollo es modular, donde diferentes equipos desarrollan diferentes módulos, conviene que cada grupo pueda disponer de las hipótesis asumidas por otros grupos para lograr una integración eficaz de todos los módulos.

Desde el punto de vista de la ejecución del modelo, conviene disponer de las hipótesis (especialmente de las implícitas) ya que en caso de que las hipótesis

dejen de ser válidas, el modelo dejará de serlo igualmente. En caso de que se realicen estudios con un modelo con hipótesis no válidas, se extraerán conclusiones erróneas.

Conviene, por último, que el registro de las hipótesis se realice a medida que se realiza el modelo, por el interés de disponer del mismo durante el desarrollo como para garantizar que registra toda la información pertinente una vez terminado el modelo.

Recogida de información correcta

Para poder comprobar la validez del modelo es necesario recoger la información pertinente de la manera correcta. Ya sea a partir de datos históricos, de tomas de tiempo ad hoc para la realización del modelo o mediante las opiniones de los expertos, conviene que la información sea la información relevante y que sea correcta. Por ejemplo, si se va a estudiar la capacidad de un determinado taller, conviene estudiar su comportamiento a partir de datos representativos de lo que se espera que sea su régimen de funcionamiento habitual y no con datos correspondientes a situaciones extremas. Por otro lado, los datos deben ser correctos; si al realizar un estudio de métodos y tiempos de los puestos de trabajo en una línea, es posible que no se obtengan los valores reales si los trabajadores de la línea son conocedores de que se está realizando dicho estudio.

Interacción con los gestores y expertos

Los gestores del sistema y los conocedores del mismo pueden, por un lado, ayudar a definir de forma precisa los objetivos del estudio si aún no están claramente definidos. Igualmente, pueden aportar su conocimiento a la hora de garantizar que el modelo representa de forma fiel el sistema real, por ejemplo, al identificar qué hipótesis son admisibles y cuáles no. Finalmente, otra de las ventajas de la interacción con los gestores es la mayor implicación de los mismos, que se traduce en una mayor credibilidad del modelo.

5. Ejecución de un modelo de simulación. Análisis de resultados

Para analizar de manera adecuada un determinado sistema de simulación, es necesario realizar de forma correcta la ejecución del modelo y el análisis de los resultados.

A veces, debido a que se interpreta la simulación como un mero ejercicio de programación, o porque se desconocen las implicaciones del carácter aleatorio del modelo, o por el coste asociado a las repeticiones, se realiza una única repetición y se toman decisiones con esta única repetición

Otros errores provienen de un tratamiento estadístico erróneo o insuficiente (realización de un número insuficiente de repeticiones, consideración de valores pertenecientes al régimen transitorio, etc.)

Conviene prestar atención a este aspecto de la simulación para evitar disponer de un modelo de simulación caro válido, del que se obtienen conclusiones equivocadas.

5.1. Tipos de simulación

Dependiendo del carácter temporal del comportamiento del sistema estudiado, se puede establecer la siguiente clasificación:

- **Simulación limitada**, propia de los sistemas en los que la duración del periodo de tiempo objeto de estudio está delimitado por algún tipo de evento. Al comienzo de este periodo, el sistema está en unas determinadas condiciones iniciales y, por lo tanto, se debe procurar que las condiciones iniciales del modelo de simulación sean representativas del sistema real. Por ejemplo, el análisis del funcionamiento de una sucursal bancaria, a lo largo de una jornada es un caso de simulación limitada.
- **Simulación ilimitada**, propia de sistemas en los que no existe un horizonte temporal determinado. A su vez, dentro de esta categoría se puede distinguir entre los siguientes casos:

- Con régimen permanente, en los que el comportamiento del sistema se estabiliza pasado un determinado tiempo. Dependiendo de las condiciones en las que comienza la simulación, se atraviesa un periodo transitorio durante el cual se obtienen valores que generalmente no son representativos del funcionamiento del sistema en condiciones normales. Si se simula una línea de montaje con todos los puestos de servicio vacíos, transcurrirá un tiempo hasta que la línea se llene y su operación sea la *'normal'*, es decir, hasta que se alcance el régimen permanente. El hecho de que se alcance el régimen permanente no significa que las variables de salida tomen un valor constante, el carácter estocástico se manifiesta igualmente en las variables de salida. Sin embargo, una vez alcanzado el régimen permanente, las variables de salida ofrecen una función de distribución constante (el patrón de comportamiento se mantiene).
- Con régimen permanente cíclico, en los que el sistema presenta un comportamiento cíclico. Si, por ejemplo, la demanda de un sistema de tipo JIT varía mensualmente, cabe esperar un régimen permanente con variaciones cíclicas que se repiten cada mes.
- Sin régimen permanente, en los que no se observa ningún tipo de patrón constante a lo largo de la simulación.

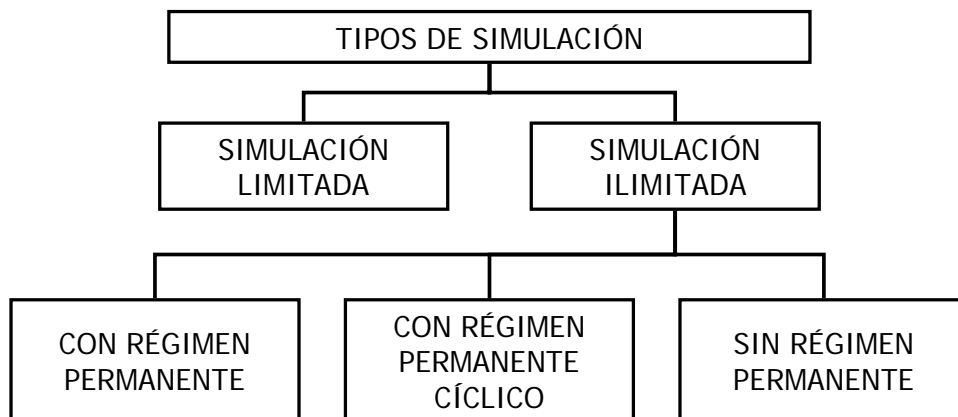


Fig.11. Clasificación de los tipos de simulación.

5.2. Análisis en simulación limitada

Selección de las condiciones iniciales

El establecimiento de las condiciones iniciales de la simulación limitada es especialmente importante, porque de ellas dependen los valores de las variables de salida a lo largo de la simulación. Se debe procurar que dichas condiciones iniciales sean representativas de las que se encuentran en el sistema real.

Por ejemplo, en el estudio de una sucursal bancaria entre mediodía y las dos de la tarde, se pueden ofrecer, por ejemplo, dos maneras de establecer las condiciones iniciales del modelo.

Una primera alternativa podría consistir en ejecutar el modelo desde la hora en la que el banco abre sus puertas, hora a la que se conoce con certeza que no existen clientes, no existen elementos en cola, que las tareas están sin comenzar, etc. Ejecutando el modelo hasta las dos de la tarde y recogiendo valores de las variables de salida sólo entre las 12 y las 2 es posible estudiar el comportamiento del sistema en las horas en la que se deseaba.

Una segunda posibilidad podría consistir en lo siguiente. Durante un número *suficiente* de días, se recogerían datos reales del sistema al mediodía (número de clientes en cola, tipo de operaciones que desean realizar, estado de las tareas, etc.) A partir de esa información sería posible generar de forma coherente, valores iniciales cada vez que se desea realizar una replicación.

Estimación de la media

Típicamente, el parámetro más interesante es la media de las variables de salida del modelo de simulación. Para el cálculo de dicho valor, a partir de los valores obtenidos en las diferentes replicaciones, se obtienen intervalos de confianza.

Es decir, si X una variable de salida (producción total diaria de una planta, por ejemplo), cuya media es $\mu = E(X)$, y $X_1, X_2 \dots X_n$, es posible construir un intervalo de confianza para la media como se explicó en el apartado 2.6 de la siguiente manera:

$$\left[\bar{X}(n) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}, \bar{X}(n) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}} \right]$$

Es posible, adicionalmente, determinar cuál es el número de replicaciones para alcanzar una precisión determinada. En concreto, el criterio que debe satisfacerse para que, con n_a^* replicaciones, el error absoluto sea menor que un determinado valor β , es el siguiente:

$$n_a^*(\beta) = \min \left\{ i \geq n, z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}} \leq \beta \right\}$$

Se puede, igualmente, establecer un criterio para el número mínimo de repeticiones de manera que el error relativo,

$$\frac{|\bar{X} - \mu|}{|\mu|}$$

sea menor que un valor dado γ :

$$n_a^*(\gamma) = \min \left\{ i \geq n, \frac{z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2(n)}{n}}}{|\bar{X}(n)|} \leq \gamma' \right\} \text{ con } \gamma' = \frac{\gamma}{1-\gamma}$$

Estimación de otros parámetros

En ocasiones puede no ser suficiente la comparación de los valores medios de una determinada variable de salida del modelo. Por ejemplo, si se realiza un estudio sobre el tiempo de ocupación de un trabajador (o un conjunto de ellos), dos configuraciones alternativas pueden arrojar el mismo resultado (o resultados similares) y que, sin embargo, se trate de configuraciones muy diferentes. Por ejemplo, podría ocurrir que, en un caso, el total de los intervalos

durante los cuales el trabajador está ocioso fueran muchos y muy breves, mientras que en el segundo podría ocurrir que el tiempo ocioso correspondiera a muy pocas paradas y de larga duración. En el segundo caso cabría la posibilidad de asignar algún tipo de tarea adicional al trabajador, mientras que en el primer caso no sería viable.

Un segundo ejemplo es el que se puede dar en un sistema de espera. Si un conjunto de puestos de servicio atiende a un conjunto de clientes que hacen cola, en un caso en una única cola y en otro en colas individuales para cada puesto de servicio, a pesar de que el tiempo medio de espera puede ser el mismo, podría ocurrir que la probabilidad de esperas muy largas no fuera la misma. Es decir, puede interesar conocer la probabilidad de que un cliente espere más de diez minutos.

Por lo anterior, puede ser interesante estudiar otros parámetros de una determinada variable. Por ejemplo, puede interesar saber cuál es la probabilidad de que una variable X tome algún valor dentro del intervalo B , es decir, $p(X \in B)$. Para ello bastaría con hacer n replicaciones independientes, con lo que se obtendría un conjunto de valores de X (X_1, X_2, \dots, X_n). Sea S el número de observaciones que caen dentro del intervalo B . La variable S es una variable binomial de parámetros p y n . La forma de estimar p (que es el parámetro que se pretende obtener con el análisis anterior), es mediante el siguiente estimador:

$$\hat{p} = \frac{S}{n}$$

Existen técnicas para calcular el cuantil q -ésimo para una determinada variable (Kelton, 1991) e intervalos de confianza para dicho cuantil (Welch, 1983)

5.3. Análisis en simulación ilimitada con régimen permanente

Cálculo del tiempo de calentamiento

En el caso de la simulación ilimitada con régimen permanente, no es crítica la selección de las condiciones iniciales, ya que aunque estas pueden condicionar la velocidad de convergencia hacia el régimen permanente, dicho régimen se acaba alcanzando en cualquier caso. El problema, en este caso consiste en la determinación del tiempo de calentamiento, que es el tiempo que debe transcurrir para que el sistema alcance el régimen permanente.

Una elección de un tiempo de calentamiento demasiado breve puede conducir a que se incorporen en el análisis valores que no son representativos del funcionamiento del sistema en régimen permanente y que la estimación de las variables de salida queden sesgadas y dependan de las condiciones iniciales.

Con un tiempo de calentamiento demasiado largo, existen garantías de que los valores son representativos. El inconveniente en este caso es la pérdida de eficiencia, ya que la duración del tiempo de ejecución superior a lo necesario.

El método más simple para la determinación del tiempo de calentamiento es de carácter gráfico (Welch, 1981 y 1983) y se basa en la construcción de medias móviles para las variables de salida.

Si en un problema de simulación ilimitada se desea estimar la media de una determinada variable Y , es decir, se desea conocer $E(Y)$, la estimación de dicho valor se hará mediante la siguiente expresión:

$$\bar{Y}(m,l) = \frac{\sum_{i=l+1}^m Y_i}{m-l}$$

Donde m es el número total de observaciones de la variable Y y l es el conjunto de las primeras observaciones que no se emplean para la estimación,

porque son la que pertenecen al régimen permanente. Si l no se calcula bien y es demasiado pequeño puede ocurrir que la discrepancia entre el valor de la media, $E(Y)$, y la estimación, \bar{Y} , sea muy importante. Si, por el contrario, se adopta la opción conservadora de optar por un valor de l muy elevado, puede reducir notablemente la eficiencia del proceso de obtención de resultados.

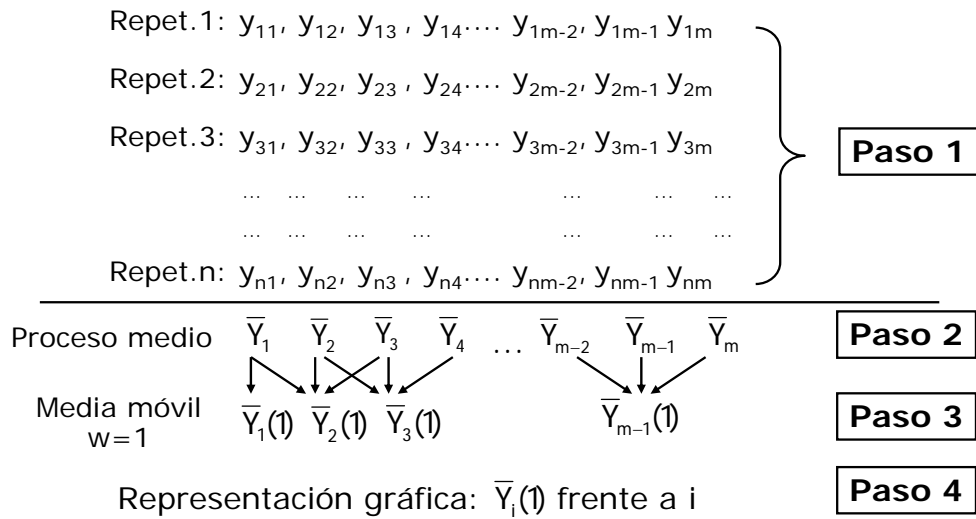


Figura X. Esquema del método de Welch para el cálculo de calentamiento

En la figura **X** se muestra un esquema del proceso. Si se realizan n repeticiones, cada una de ellas de una longitud m y se toman m observaciones de la variable Y en cada repetición, se define y_{ij} como el valor de la observación de la variable Y en el instante j de la repetición i . Primero, se calculan los valores \bar{Y}_j , que representa la media de las medidas de Y para las n repeticiones para el instante j , es decir:

$$\bar{Y}_j = \frac{\sum_{i=1}^n y_{ij}}{n}$$

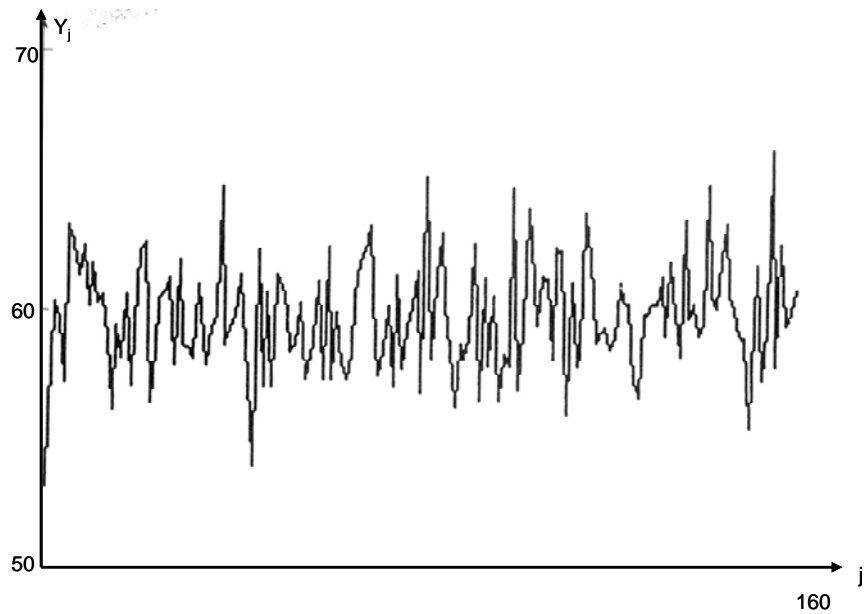


Figura X. Representación de \bar{Y}_i

A continuación se calcula la media móvil de la siguiente manera:

$$\bar{Y}_i(w) = \begin{cases} \frac{\sum_{s=-w}^w \bar{Y}_{i+s}}{2w+1}, & \text{si } i = w+1, \dots, m-w \\ \frac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1}, & \text{si } i = 1, \dots, w \end{cases}$$

Donde w es la amplitud de la media móvil. Inicialmente, w toma un valor pequeño, por ejemplo 2. Una vez que se han calculado los valores de $\bar{Y}_i(w)$ se representan gráficamente. Como se trata de medias móviles, la representación gráfica será más 'suave' que la representación de los valores \bar{Y}_j . Es muy probable que para valores pequeños de w la representación de $\bar{Y}_i(w)$ no se estabilizará en ningún valor, con lo cual será necesario calcular medias móviles tomando valores mayores de w , hasta que se aprecie, como en la figura **X**, que los valores de $\bar{Y}_i(w)$ se estabilizan en torno a un valor a partir de un

determinado l , que es, precisamente el tiempo de calentamiento (en este caso, 24).

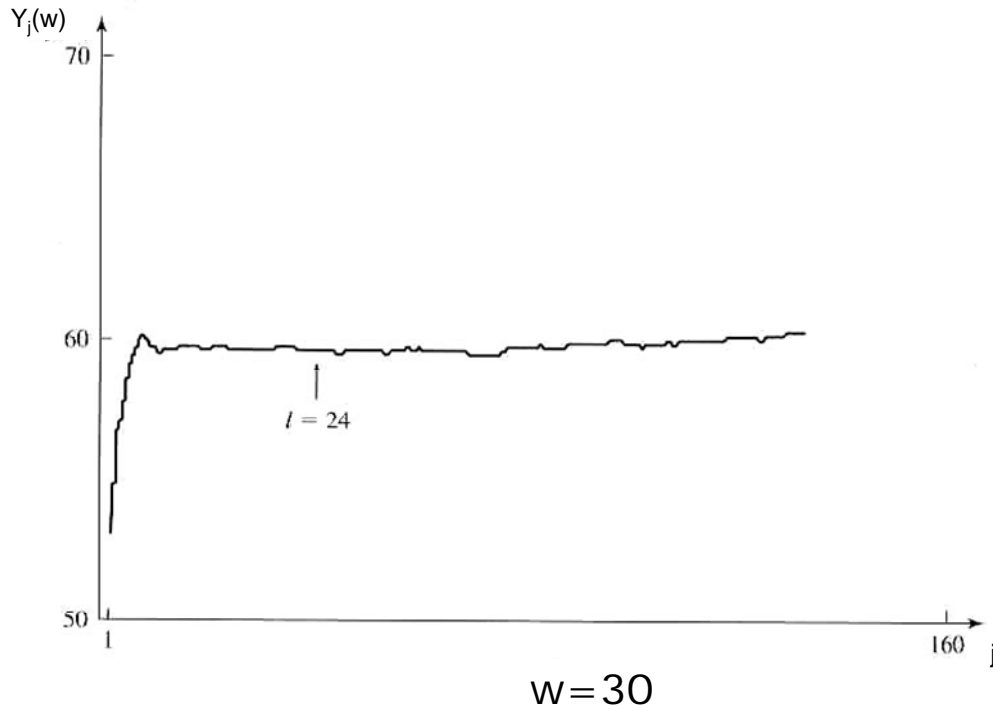


Figura X. Representación de $\bar{Y}_i(w)$

•

Por último, conviene apuntar algunas recomendaciones para la selección de los parámetros:

- n (nº de repeticiones): entre 5 y 10 inicialmente. Si no se consigue un w razonable hacer otras tantas repeticiones
- m (longitud de las repeticiones): grande en comparación con el valor previsto de l y razonablemente grande para que ocurran varias veces sucesos infrecuentes.
- w (parámetro de amplitud de la media móvil): probar valores hasta obtener el mínimo w que hace la convergencia razonablemente suave

Estimación de parámetros

Existen diferentes formas de realizar las replications para estimar los parámetros de las variables de salida. A continuación se comentan dos.

- Múltiples replications. Una primera alternativa consiste en la repetición de n replications de una longitud m . Como se ha indicado en el punto anterior, es necesario dejar transcurrir el tiempo de calentamiento antes de comenzar a recoger valores de las variables de salida. Con los valores de cada una de las diferentes replications, se realiza un análisis análogo al que se ha descrito para el caso de la simulación limitada.

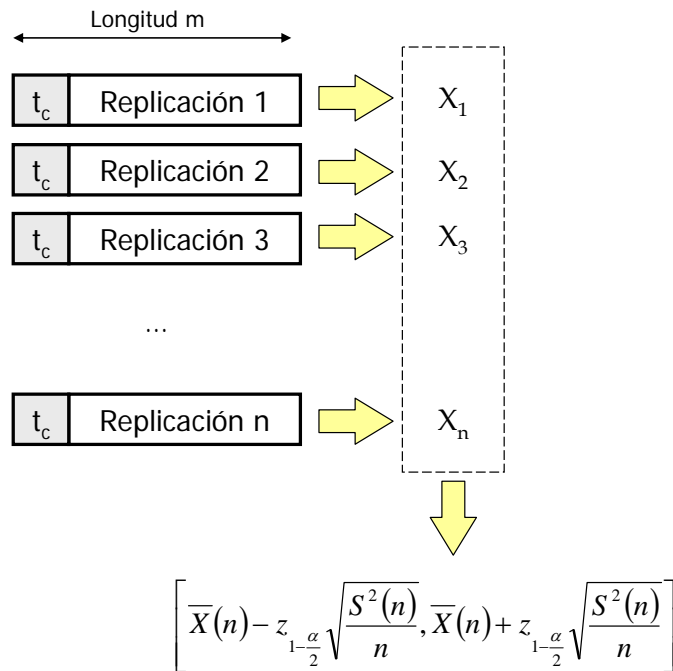


Fig. 12. Análisis en simulación ilimitada con régimen permanente. Múltiples replications

- Única replicación *larga*. En este caso se realiza una única replicación larga de longitud l y, una vez finalizada, se fracciona dicha replicación en n cuya longitud es:

$$m = \frac{l - t_c}{n}$$

donde l_c es el tiempo de calentamiento. Esta alternativa, tiene la ventaja de que sólo se simula el régimen transitorio una sola vez. Sin embargo, todos los valores están correlacionados, lo cuál representa un problema para realizar el análisis propuesto en los casos anteriores. Sin embargo, si la longitud de las replicaciones es suficientemente larga, el problema de la correlación desaparece.

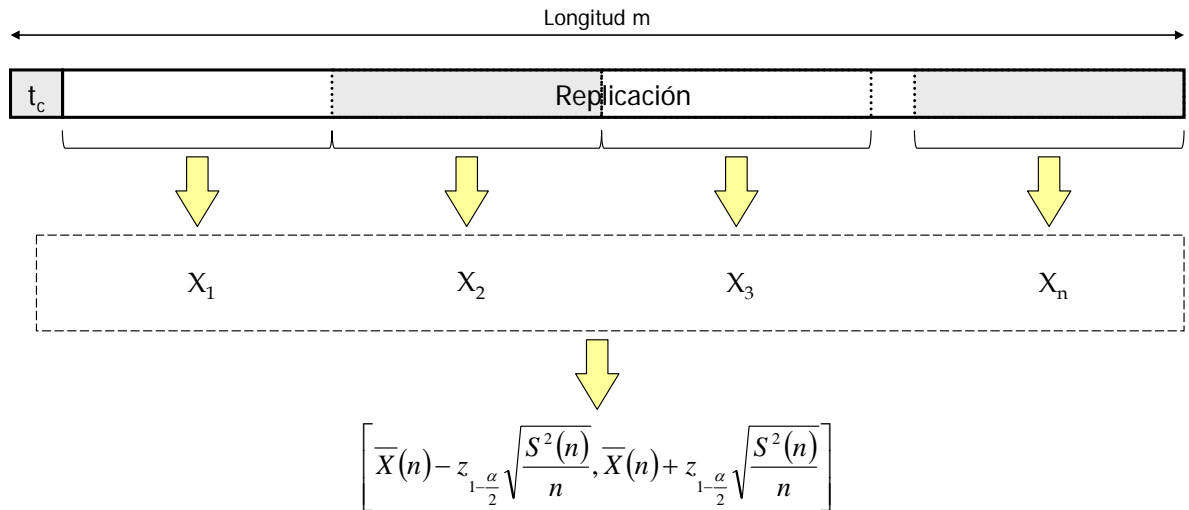


Fig. 13. Análisis en simulación ilimitada con régimen permanente. Única replicación *larga*

6. Referencias y bibliografía

Law y Kelton (1991)

Estadística

Fishman, G. S. (1971). *Principles of Discrete Event Simulation*. John Wiley. Nueva York.

Knuth, D. E. (1973). *The Art of Computer Programming. Vol. 3: Sorting and Searching*. Addison-Wiley. Reading, Mass.

Manuales de software de simulación