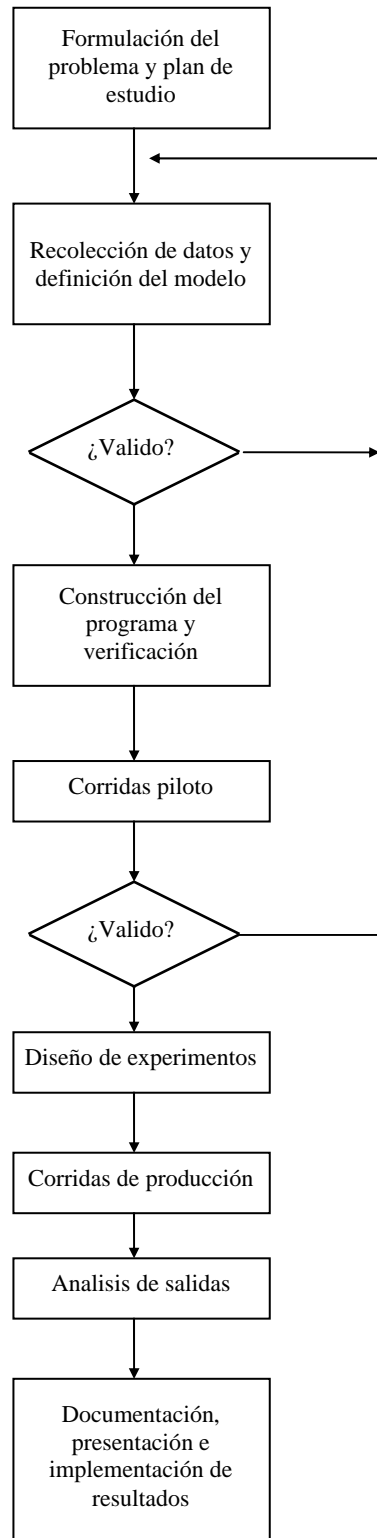


PASOS EN UN ESTUDIO DE SIMULACIÓN



Después de haber visto con detalles el interior de un simulador por eventos discretos, es bueno reconocer que codificación y modelado detallados son solo una parte del esfuerzo total y que también debemos prestar

atención a otros aspectos que van desde diseño estadístico de experimentos hasta manejo de personal y presupuesto.

El gráfico muestra los pasos que componen un estudio de simulación típico. No todos los estudios deben contener necesariamente todos los pasos y en el orden dado. Un estudio tampoco es un proceso secuencial y frecuentemente se regresa a pasos anteriores. A continuación se explican los pasos.

Formulación del problema y plan de estudio. Todo estudio debe comenzar con una muy clara definición de objetivos, de detalles específicos que se quieren cubrir, de alternativas de diseño y de criterios para evaluarlas. Hay que planificar el estudio en términos de personal, costo y tiempo.

Recolección de datos y definición del modelo. Información y datos deben ser recolectados del sistema de interés (si existe) para ser usados en la determinación de aspectos operativos y de las distribuciones de las variables aleatorias usadas en el modelo. La construcción del modelo aun es en buena medida arte. Se debe comenzar con un modelo sencillo que pueda ser refinado de ser necesario.

¿Valido? La validación debe efectuarse durante todo el estudio, sin embargo hay ciertos puntos en el estudio en que es particularmente apropiada. Los modeladores deben interactuar estrechamente con las personas conocedoras del sistema y con los tomadores de decisiones. Esto incrementa la validez y la credibilidad del modelo. También deben usarse técnicas estadísticas para verificar las distribuciones de las variables aleatorias usadas.

Construcción del programa y verificación. Hay que decidir que lenguaje usar. Hay que asegurarse de tener generadores de números aleatorios adecuados y de generar correctamente las distribuciones necesarias (estos tópicos serán cubiertos mas adelante en el curso). Hay que verificar o depurar el programa y esto se discute mas adelante.

Corridas piloto. Estas corridas se hacen para validar el modelo verificado y lo veremos con mas detalle mas adelante.

¿Valido? Las corridas piloto pueden ser usadas para determinar la sensibilidad del modelo a pequeños cambios en los parámetros de entrada. Cambios importantes implican que una mejor estimación de estos parámetros debe ser obtenida.

Diseño de experimentos. Hay que decidir que diseños del sistema se simularan y en cada caso hay que decidir sobre el tiempo de corrida de la simulación, las condiciones iniciales, la longitud del estado transitorio, y el número de replicaciones.

Corridas de producción. Son para producir datos para medir el desempeño de los diseños de interés.

Análisis de salidas. Técnicas estadísticas son usadas para analizar las salidas de las corridas de producción. Usualmente se construyen intervalos de confianza de alguna medida de desempeño de un diseño o para decidir cual es mejor respecto a cierta medida de desempeño.

Documentación, presentación e implementación de los resultados. Es importante documentar los supuestos realizados así como el programa mismo. Muchisimas veces un modelo es usado para varias aplicaciones. Un estudio de simulación cuyos resultados no sean implementados usualmente es un fracaso.

Durante el desarrollo del modelo de simulación hay que asegurarse de que el modelo sea correctamente implementado y que sea representativo del sistema real. Estos dos pasos se denominan verificación y

validación del modelo respectivamente.

I. TÉCNICAS DE VERIFICACIÓN DEL MODELO

Verificación también puede ser llamado depuración, esto es, asegurarse de que el modelo hace lo que debe hacer. En la literatura pueden encontrarse diversas técnicas de depuración. Cualquier combinación de ellas puede ser usada para verificar el modelo.

1. Diseño modular de arriaba-abajo

Los modelos de simulación resultan en programas grandes. Toda técnica que ayude a desarrollar, depurar y mantener programas grandes también es útil en simulación. Dos técnicas importantes son modularidad y diseño de arriaba-abajo.

Modularidad

El modelo debe ser estructurado en módulos que se comunican por interfaces bien definidas (variables de entrada y salida). Usualmente se llaman subrutinas, funciones, procedimientos, etc. Al estar bien definida la interface y la función del modulo, este puede ser desarrollado, depurado y mantenido independientemente. Esto permite dividir problema de verificación en problemas más pequeños.

Diseño de arriaba-abajo

Consiste en desarrollar una estructura jerárquica del modelo de forma que el problema es dividido recursivamente en problemas más pequeños. Inicialmente se divide el modelo en un número de módulos. Estos módulos a su vez son subdivididos en otros módulos. Se continúa hasta tener módulos de fácil depuración y mantenimiento.

2. Depuración

Consiste en incluir chequeos y salidas adicionales que permitan detectar errores.

3. Explicación de los módulos

Consiste en explicar el código a otros miembros del grupo. Al explicar cuidadosamente el código tanto el autor como los oyentes puede encontrar errores.

4. Modelos determinísticos

Dado que puede ser muy complicado depurar programas de modelos con variables aleatorias, las distribuciones se pueden hacer constantes (determinísticas) para facilitar la determinación de las salidas y encontrar errores.

5. Correr casos simplificados

El modelo puede ser corrido con casos sencillos. Por ejemplo, en la simulación de una red se puede usar solo un nodo intermedio y enviar solo un paquete. Que el modelo funcione con casos sencillos no es garantía de que funcione con casos más complejos. Por lo tanto, los casos deben ser tan complicados como sea posible analizarlos de forma fácil sin simulación.

6. Trazado

Un trazado es una lista de eventos ordenada por tiempo y sus variables asociadas que permite observar como se esta comportando la simulación. Debido a que los trazados representan un costo de procesamiento adicional, se deben incluir como una opción que se pueda activar o desactivar. El trazado se puede hacer a diferentes niveles de detalle y el usuario debe poder seleccionar el nivel que desee. Por ejemplo, pude seleccionar un trazado de solo aquellos eventos que pertenezcan a un nodo particular en la simulación de una red.

7. Gráficas en línea

Gráficas pueden presentar la misma información de un trazado pero en forma más comprensiva.

8. Pruebas de continuidad

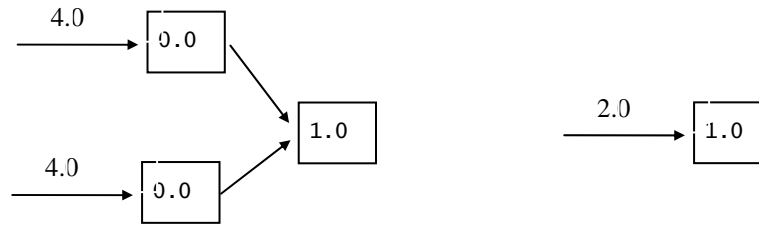
Consiste en correr la simulación varias veces con ligeras variaciones en los parámetros de entrada. Para un parámetro dado, un pequeño cambio en su valor inicial generalmente produce solo pequeños cambios en la salida. Un cambio brusco en la salida debe ser investigado.

9. Pruebas de degeneración

Consiste en chequear el modelo usando valores extremos. A pesar de que estos valores extremos no representen casos típicos, pueden ayudar a encontrar errores en los cuales el analista no hubiera pensado de otra forma. Deben incluirse chequeos de valores de entrada para verificar que estén en los rangos permitidos y de que el modelo funcione para todas las combinaciones en los rangos permitidos.

10. Pruebas de consistencia

Consiste en chequear que el modelo produce resultados similares para entradas que tengan efectos similares. Por ejemplo, en una red, dos fuentes con una tasa de llegada de 100 paquetes por segundo deben producir una carga similar a cuatro fuentes con una tasa de llegada de 50 paquetes por segundo. Si hay diferencia, esta debe ser explicada o puede ser debido a errores. Por ejemplo, el simulador de redes visto en clase debe producir resultados similares para las dos estructuras siguientes:



11. Independencia de semillas

Las semillas usadas no deben afectar las conclusiones finales. Se debe verificar que los resultados sean similares para diferentes semillas.

II. TÉCNICAS DE VALIDACIÓN DEL MODELO

Validar es asegurarse de que los supuestos usados en el desarrollo del modelo son razonables en el sentido de que, si correctamente implementado, el modelo producirá resultados próximos a los observados en el sistema real. Las técnicas de validación dependen de los supuestos hechos y del sistema modelado. Técnicas usadas en una simulación pueden no ser aplicables a otra.

Hay que validar tres puntos claves del modelo:

- a) Supuestos
- b) Valores de los parámetros de entrada y las distribuciones
- c) Valores de salida y conclusiones

Esto se puede hacer comparando con las tres posibles fuentes siguientes:

- a) Intuición de expertos
- b) Mediciones en el sistema real
- c) Resultados teóricos

No todas estas fuentes son posibles en todos los casos.

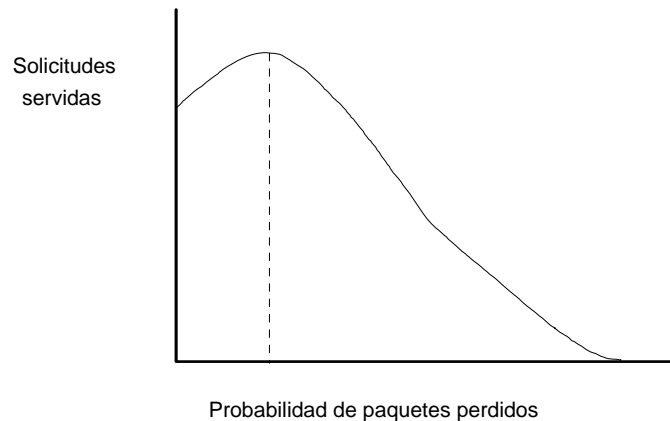
1. Intuición de expertos

Es la forma más práctica y común de validar modelos. Personas muy conocedoras del sistema son reunidas y los puntos claves son expuestos y discutidos. Es recomendable validar los puntos por separado y a medida que se desarrolla el modelo y no esperar al final.

Los supuestos son los primeros en discutirse y al tener un modelo preliminar. Las entradas al ir desarrollando el modelo. Finalmente las salidas se validan al tener un modelo ejecutable que ha sido verificado.

Ejemplo

Supongamos que en la simulación de una red se obtiene el siguiente gráfico que relaciona el número de solicitudes procesadas en función de paquetes perdidos. Este gráfico es claramente contraintuitivo ya que sugiere que el sistema se comporta mejor con cierta probabilidad de pérdidas que sin ellas.



2. Mediciones en el sistema real

Comparar con el sistema real es la forma más segura y preferible de validar el modelo. En la práctica esto no siempre es posible ya que el sistema no existe o es muy costoso obtener mediciones. Técnicas estadísticas pueden ser usadas para comparar si dos sistemas son iguales. Pruebas de bondad de ajuste pueden ser usadas para las distribuciones de entrada.

3. Resultados teóricos

A veces es posible modelar analíticamente el sistema usando supuestos simplificados. También puede ser posible determinar las distribuciones de entrada analíticamente. En estos casos los resultados teóricos y los de la simulación pueden ser usados para validar el modelo. Hay que tener cuidado ya que ambos modelos pueden ser inválidos debido a los supuestos y simplificaciones que tienen.

“Validar completamente un modelo es imposible”. Solo es posible demostrar que no es inválido para ciertas situaciones. La validación esta limitada a un reducido numero de escenarios y la intención es cubrir los casos más relevantes.

III. SUPRESIÓN DEL ESTADO TRANSITORIO

En muchos estudios de simulación se invierte mucho tiempo y dinero en el desarrollo del modelo y en la programación, pero poca atención es prestada al análisis de los resultados. Un procedimiento normal consiste

en tomar las estimaciones resultantes de **una** simulación como las verdaderas características del sistema. Dado que en general se usan muestras aleatorias durante la simulación, estos estimados son solo una realización de variables aleatorias que pueden tener varianzas muy grandes. Por lo tanto los estimados de una corrida pueden diferir significativamente de las características respectivas del modelo y las conclusiones derivadas de ellos pueden ser erróneas.

En la mayoría de las simulaciones solo son de interés los resultados del sistema cuando este está en un **estado estable**. En estos casos, resultados de la parte inicial de la simulación no deben ser incluidos en las conclusiones finales. Esta parte inicial es llamada el **estado transitorio** y debe ser removido. El mayor problema de hacer esto es que no es posible definir con exactitud lo que constituye el estado transitorio y cuando este termina. Por lo tanto, todos los métodos de supresión del estado transitorio son heurísticos y entre ellos tenemos:

1. Corridos prolongados
2. Inicialización adecuada
3. Truncamiento
4. Eliminación de datos iniciales
5. Traslado de medias de repeticiones independientes
6. Medias de tandas

1. Corridos prolongados

Implica hacer los corridos lo suficientemente prolongados para hacer que el efecto de las condiciones iniciales sea despreciable en los resultados finales.

Tiene dos desventajas fundamentales:

Primero, desperdicia recursos y si estos son costosos las simulaciones no deben prolongarse más de lo estrictamente necesario.

Segundo, es difícil asegurar que la simulación ha sido lo suficientemente larga.

Por estas dos razones, este método debe evitarse.

2. Inicialización adecuada

Requiere iniciar la simulación en un estado próximo al estado estable. Esto resulta en una reducción del estado transitorio y de su efecto en las conclusiones finales.

3. Truncamiento

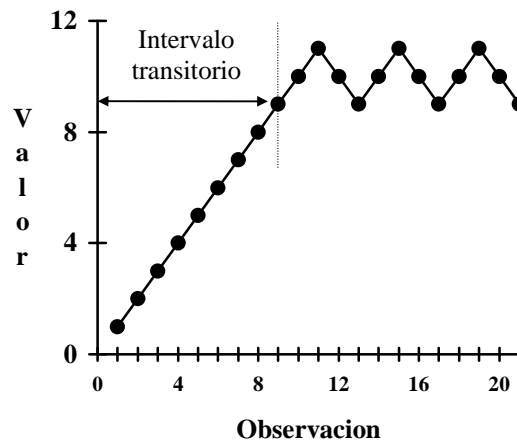
Este método y los siguientes se basan en el supuesto de que la variabilidad durante el estado transitorio es mayor que durante el estado estable, que en general es cierto.

En el método de truncamiento, la variabilidad es medida en términos de rangos: las observaciones mínimas y máximas. Si observaciones sucesivas son graficadas, se puede ver que las observaciones tienden a estabilizarse cuando la simulación entra en estado estable.

Dada una muestra de n observaciones $\{x_1, x_2, x_3, \dots, x_n\}$, el método de truncamiento consiste en ignorar las primeras l observaciones y calcular el mínimo y máximo de las restantes $n - l$ observaciones. Esto es repetido para $l = 1, 2, 3, \dots, n - 1$ hasta que la observación $l + 1$ no sea ni el mínimo ni el máximo de las restantes observaciones. Este valor de l da la longitud del estado transitorio.

Ejemplo

Consideremos la siguiente secuencia de observaciones: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 10, 9, 10, 11, 10, 9, 10, 11, 10, 9, ... Ignorando la primera observación ($l = 1$), el rango restante es (2, 11). Dado que la segunda observación es el mínimo, la fase transitoria es mayor que 1. Continuamos de esa forma hasta tener $l = 9$ en donde el rango de la secuencia restante es (9, 11) y la décima observación no es ni el mínimo ni el máximo. Por lo tanto la longitud del intervalo transitorio es 9, y las primeras nueve observaciones son descartadas.



Este método puede dar resultados incorrectos. Consideremos la siguiente secuencia: 11, 4, 2, 6, 5, 7, 10, 9, 10, 9, 10, 9, 10, ... El método daría un intervalo transitorio incorrecto de 1 ya que 4 no es ni el mínimo ni el máximo de las observaciones restantes.

4. Eliminación de datos iniciales

Este método requiere estudiar las medias globales después de que ciertas observaciones iniciales han sido eliminadas de la muestra. En estado estable, las medias no cambian significativamente (hay cambios por aleatoriedad) a lo que observaciones son eliminadas. Para reducir el efecto aleatorio, el método requiere medias de varias replicaciones. Cada replicación es una corrida completa con los mismos parámetros iniciales y difieren solo en las semillas usadas. Promediando sobre las replicaciones tendremos una secuencia más suave.

Supongamos m replicaciones de tamaño n . Sea x_{ij} la j -ésima observación de la i -ésima replicación. El método consta de los siguientes pasos:

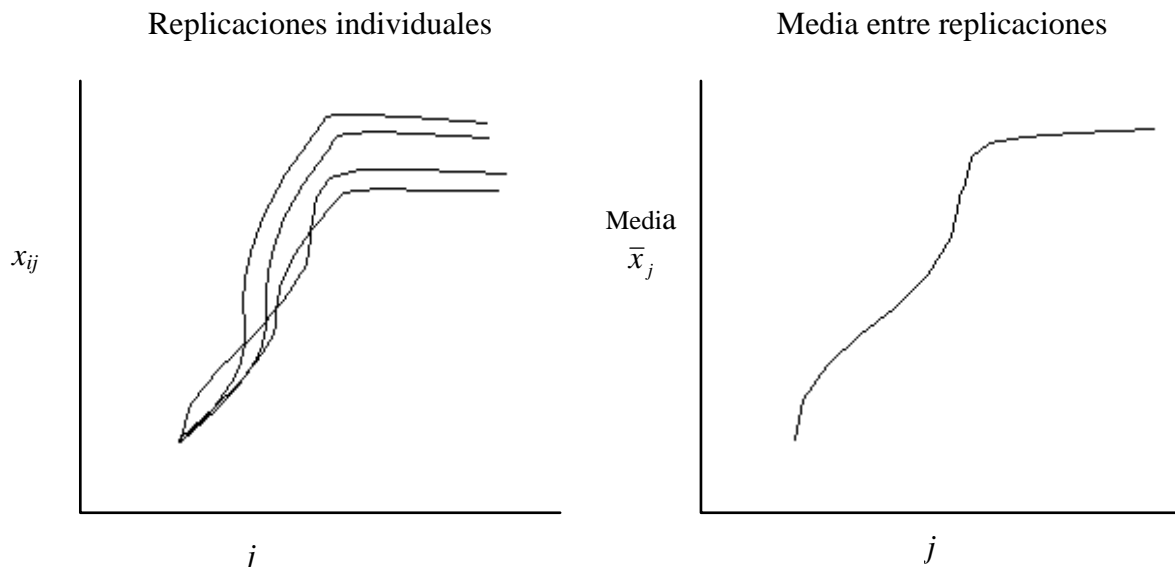
1. Obtengamos la trayectoria promedio entre las replicas:

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, \quad j = 1, 2, \dots, n$$

2. Obtengamos la media global:

$$\bar{\bar{x}} = \frac{1}{n} \sum_{j=1}^n \bar{x}_j$$

Fijemos $l = 1$ y vayamos al siguiente paso.



3. Asumiendo que el estado transitorio es de longitud l , eliminemos las primeras l observaciones de la trayectoria promedio y obtengamos la media global de los restantes $n - l$ valores:

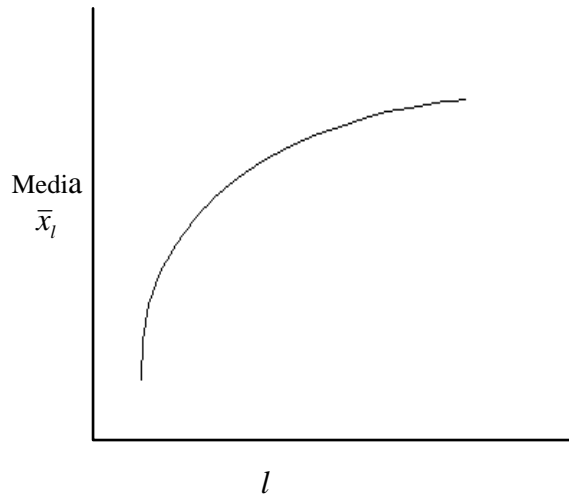
$$\bar{\bar{x}}_l = \frac{1}{n-l} \sum_{j=l+1}^n \bar{x}_j$$

4. Calculemos el cambio relativo en la media global:

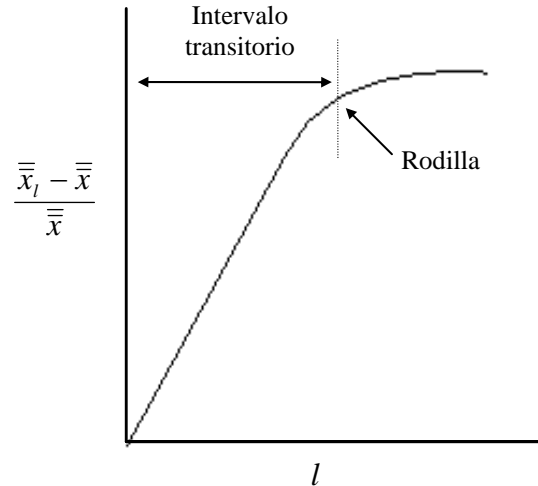
$$\text{Cambio relativo} = \frac{\bar{\bar{x}}_l - \bar{\bar{x}}}{\bar{\bar{x}}}$$

5. Repitamos los pasos 3 y 4 variando l desde 1 hasta $n - 1$. Después de un cierto valor de l , la gráfica del cambio relativo se estabiliza. Este valor es conocido como la rodilla y representa la longitud del intervalo transitorio.

Media de las últimas $n-l$ observaciones



Cambio relativo



5. Traslado de medias de replicaciones independientes

Este método es similar al anterior. La diferencia principal es que este método requiere calcular las medias sobre una ventana de tiempo movable en vez de medias globales.

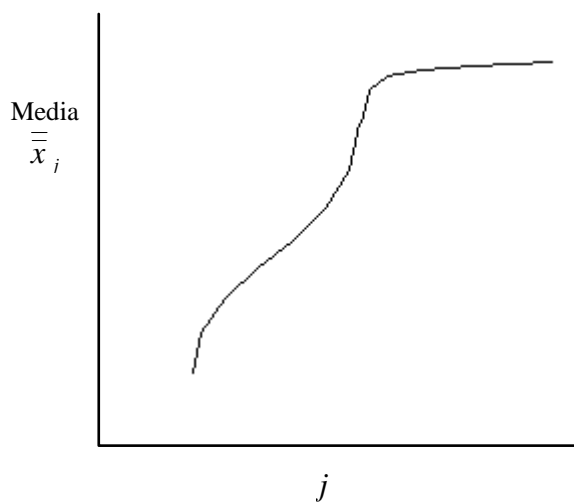
Supongamos nuevamente m replicaciones de tamaño n . Sea x_{ij} la j -ésima observación de la i -ésima replicación. El método consta de los siguientes pasos:

1. Obtenemos la trayectoria promedio entre las replicaciones:

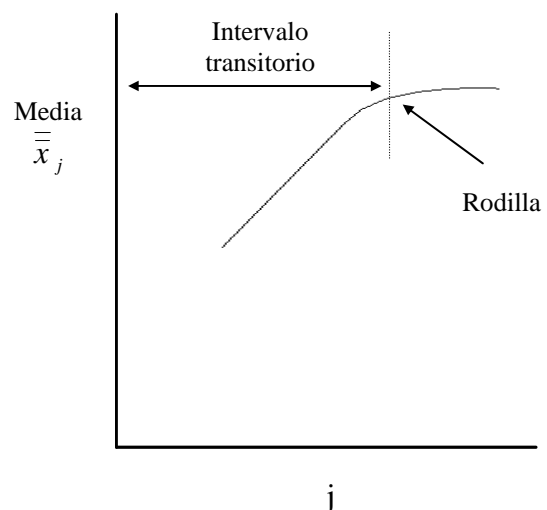
$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, \quad j = 1, 2, \dots, n$$

Fijemos $k = 1$ para el siguiente paso.

Media móvil con $k = 1$



Media móvil con $k = 5$



- Grafiquemos la trayectoria de los promedios móviles de los sucesivos $2k + 1$ valores:

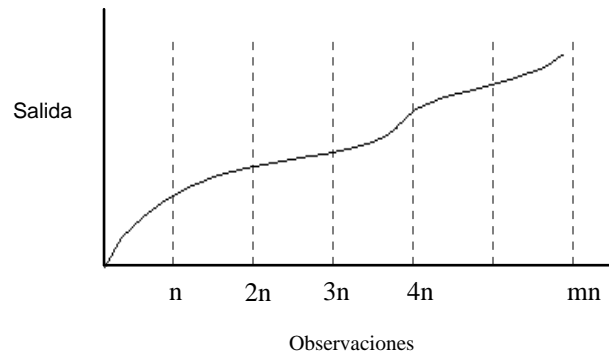
$$\bar{\bar{x}}_j = \frac{1}{2k + 1} \sum_{l=-k}^k \bar{x}_{j+l}, \quad j = k + 1, k + 2, \dots, n - k$$

- Repitamos el paso 2 con $k = 2, 3, \dots$ hasta que la gráfica sea lo suficientemente suave.
- Encontremos la rodilla en el gráfico. El valor de j da la longitud de la fase transitoria.

6. Medias de tandas

Este método requiere una simulación prolongada que luego será dividida en varias partes de igual duración. Cada parte es llamada una **tanda** o **submuestra**. La media de las observaciones en una tanda es llamada **media de la tanda**. El método estudia la varianza de estas medias como función del tamaño de la tanda.

Como se muestra en la figura, una corrida de N observaciones puede ser dividida en m tandas de tamaño n , donde $m = \lfloor N/n \rfloor$. Aquí $\lfloor \cdot \rfloor$ significa el “máximo entero menor o igual a”. Sea x_{ij} la j -ésima observación de la i -ésima tanda.



- Para cada tanda calculemos la media:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}, \quad i = 1, 2, \dots, m$$

- Calculemos la media global:

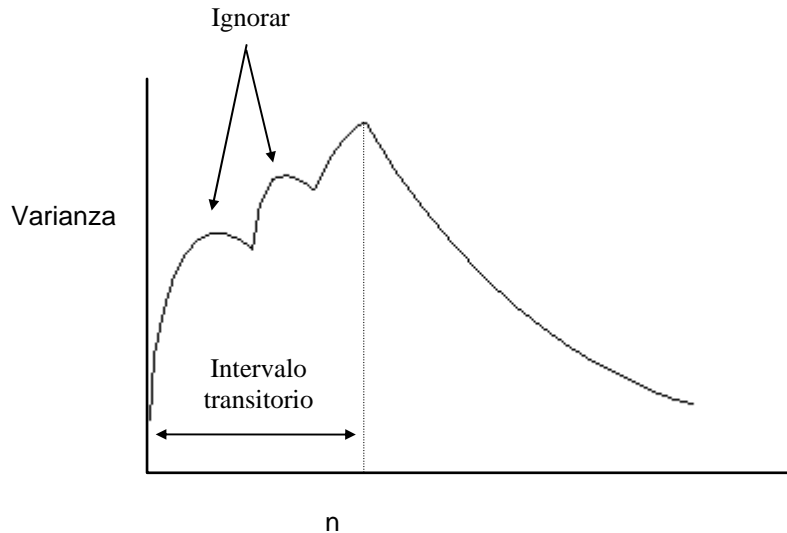
$$\bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

- Calculemos la varianza de las medias de las tandas:

$$Var(\bar{x}) = \frac{1}{m-1} \sum_{i=1}^m (\bar{x}_i - \bar{\bar{x}})^2$$

Incrementemos el valor de n y repitamos los pasos 1 y 3 para $n = 3, 4, 5, \dots$. Grafiquemos la varianza en función de n y la longitud del intervalo transitorio es aquel en donde la varianza definitivamente comienza a decrecer.

La idea es que si la longitud de la fase transitoria es T entonces, si el tamaño de las tandas n es mucho menor que T , las tandas iniciales hacen que la media global se aproxima a las medias de las tandas iniciales y la varianza es pequeña. A lo que el tamaño de las tandas crece, la varianza se incrementa. Cuando n es más grande que T , solo la media de la primera tanda es diferente y las restantes son aproximadamente iguales. Esto hace que la varianza decrezca.



III.a. SIMULACIONES TERMINANTES

A pesar de que en la mayoría de las simulaciones lo que interesa es el comportamiento estado estable, hay sistemas en los que nunca se alcanza este estado. Estos sistemas siempre operan en estado transitorio. En estos sistemas hay un evento natural E que determina el tiempo de simulación de cada corrida o replicación. En estos casos es necesario estudiar el sistema en estado transitorio. Estas son llamadas **simulaciones terminantes**. Algunos ejemplos son:

- a) Una oficina/comercio que trabaja cierto horario. Si un comercio abre a las 8:00 a.m. y cierra a las 6:00 p.m. entonces $E = \{\text{al menos 10 horas han pasado y no hay clientes en el comercio}\}$.
- b) Una confrontación militar que termina cuando algún bando ha perdido 30% de su fuerza. $E = \{\text{uno de los bandos gana la confrontación}\}$.
- c) Una empresa recibe una orden para construir 1000 cavas para cierto tipo de camiones. Se quieren simular distintas opciones de ensamblado. $E = \{\text{las 1000 cavas han sido completadas}\}$.

En este tipo de simulaciones por lo general se realizan varias replicaciones para cada iteración y se sacan conclusiones usando promedios sobre las replicaciones.

Ejemplo:

Consideremos un banco con 5 taquillas y una sola cola que abre a las 9 a.m. y cierra a las 5 p.m., pero atiende a todos los clientes en el banco a las 5 p.m. Los clientes llegan de acuerdo a una distribución de Poisson a razón de 1 por minuto, el tiempo de servicio es exponencial con media 4 minutos y los clientes son atendidos por orden de llegada. La tabla siguiente muestra los resultados de 10 replicaciones independientes.

Replicacion	Clientes servidos	Tiempo total de servicio (h)	Tiempo promedio de espera (min)	Longitud de cola promedio	Proporcion de clientes retrasados < 5 min
1	484	8.12	1.53	1.52	0.917
2	475	8.14	1.66	1.62	0.916
3	484	8.19	1.24	1.23	0.952
4	483	8.03	2.34	2.34	0.822
5	455	8.03	2.00	1.89	0.840
6	461	8.32	1.69	1.56	0.866
7	451	8.09	2.69	2.50	0.783
8	486	8.19	2.86	2.83	0.782
9	502	8.15	1.70	1.74	0.873
10	475	8.24	2.60	2.50	0.779
Media	475.60	8.15	2.03	1.97	0.853
Varianza	249.38	0.01	0.31	0.28	0.004

Una estimación del tiempo promedio de espera esta dado por:

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 2.03$$

y un intervalo de confianza de aproximadamente el 90% es:

$$\bar{x} \pm t_{1-\alpha/2; n-1} \sqrt{\frac{\text{Var}(x)}{n}} = 2.03 \pm t_{0.95; 9} \sqrt{\frac{0.31}{10}} = 2.03 \pm 1.833 \times 0.176 = 2.03 \pm 0.32$$

y podemos decir que con un 90% de confiabilidad el tiempo promedio de espera esta contenido en el intervalo [1.71,2.35] minutos.

Un aspecto relacionado es el de las **condiciones finales** -- las condiciones al final de la simulación. El estado del sistema al final de la simulación puede no ser típico de un estado estable. En estos casos es necesario excluir la porción final de la salida. Métodos para esto son similares a los vistos para remover el estado transitorio.

Finalmente hay que tener cuidado con los resultados al final de la simulación. Por ejemplo, consideremos un procesador que ejecuta trabajos. Para calcular los tiempos promedios de servicio solo debemos incluir aquellos trabajos que fueron servidos (no hay que incluir los que están siendo servidos):

$$\text{Tiempos de servicio promedio} = \frac{\text{tiempo total de servicio}}{\text{numero de trabajos que fueron servidos}}$$

Para calcular el tiempo promedio de espera, solo aquellos trabajos que completaron la espera y comenzaron servicio deben ser incluidos:

$$\text{Tiempos promedio de espera} = \frac{\text{suma de los tiempos de espera}}{\text{numero de trabajos que comensaron servicio}}$$

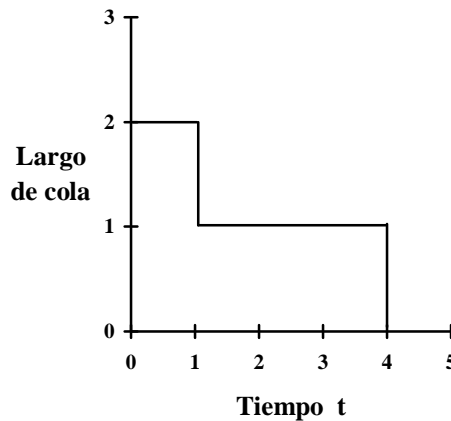
Si q_j es la longitud de la cola al j -esimo evento que resulto en un cambio de longitud de cola, el promedio de los q_j no da la longitud promedio de cola:

$$\text{Longitud promedio de cola} \neq \frac{\sum_{j=1}^n (\text{longitud de cola al evento } j)}{\text{numero de eventos } n}$$

En vez, un promedio de tiempo de la longitud de cola debe ser usado:

$$\text{Longitud promedio de cola} = \frac{1}{T} \int_0^T \text{Longitud de cola}(t) dt$$

Ejemplo:



La figura muestra las longitudes de una cola. En el tiempo $t = 0$ llegan dos trabajos. Uno de los trabajos finaliza al tiempo $t = 1$ y el segundo parte al tiempo $t = 4$. Hay tres eventos: una llegada y dos salidas. Las longitudes de cola son 2, 1, y 0, pero la longitud promedio de la cola no es $(2 + 1 + 0) / 3 = 1$, sino $(2 \times 1 + 1 \times 3) / 4 = 5/4 = 1.25$.

IV. CRITERIOS DE TERMINACIÓN

Es muy importante que el tiempo de simulación sea adecuado. Si la simulación es muy corta, los resultados pueden ser altamente variables. Si por el contrario es muy larga, recursos humanos y de computación pueden ser desperdiciados. La simulación debe correr hasta que el intervalo de confianza de la media de las respuestas se aproxime a la amplitud deseada. Si la media muestral es \bar{x} y su varianza es $Var(\bar{x})$, un intervalo de confianza de $100(1-\alpha)\%$ para la media esta dado por:

$$\bar{x} \pm z_{1-\alpha/2} \sqrt{Var(\bar{x})}$$

donde $z_{1-\alpha/2}$ es el $(1-\alpha/2)$ -esimo cuantil de una normal estándar $N(0,1)$. La varianza de una media muestral de n observaciones *independientes* se puede obtener de la varianza de las observaciones:

$$Var(\bar{x}) = \frac{Var(x)}{n}$$

Nota:

Esto sale de la aplicación del Teorema Central del Límite para la estimación de la media poblacional a partir de la media muestral:

$$P(\bar{x} - z_{1-\alpha/2}s / \sqrt{n} \leq \mu \leq \bar{x} + z_{1-\alpha/2}s / \sqrt{n}) = 1 - \alpha$$

donde:

$(\bar{x} - z_{1-\alpha/2}s / \sqrt{n}, \bar{x} + z_{1-\alpha/2}s / \sqrt{n})$	es el intervalo de confianza ,
α	es el nivel de significación ,
$100(1-\alpha)$	es el nivel de confianza , y
$(1-\alpha)$	es el coeficiente de confianza .

Esta formula se puede usar solo si las observaciones son independientes. Desafortunadamente este no es el caso en la mayoría de las simulaciones. Por ejemplo, en la simulación de colas, si el tiempo de espera del i -esimo trabajo es grande, el tiempo de espera del $(i + 1)$ -esimo trabajo también será grande y viceversa. En este caso los tiempos de espera sucesivos están altamente correlacionados y la formula anterior no se puede usar. Cuando las observaciones están correlacionadas se puede usar uno de los siguientes métodos:

- Replicaciones independientes
- Medias de tandas
- Regeneración

1. Replicaciones independientes

Este método se basa en el supuesto de que las medias de replicaciones independientes son independientes a pesar de que las observaciones en una replicación particular son correlacionadas.

El método consiste en realizar m replicaciones independientes de tamaño $n + n_0$ cada una, donde n_0 es el

tamaño de la fase transitoria y la cual es descartada. Luego se computa lo siguiente:

1. Calculemos la media de cada replicación:

$$\bar{x}_i = \frac{1}{n} \sum_{j=n_0+1}^{n_0+n} x_{ij}, \quad i = 1, 2, \dots, m$$

2. Calculemos la media global sobre todas las replicaciones:

$$\bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

3. Calculemos la varianza de la media de las replicaciones:

$$Var(\bar{x}) = \frac{1}{m-1} \sum_{i=1}^m (\bar{x}_i - \bar{\bar{x}})^2$$

El intervalo de confianza para la media es

$$\left[\bar{\bar{x}} \pm z_{1-\alpha/2} \sqrt{Var(\bar{x})} \right] \quad \text{Si } m < 30 \text{ se debe usar } t_{[1-\alpha/2; m-1]} \text{ en vez de } z_{1-\alpha/2}.$$

Nótese que el método requiere descartar mn_0 observaciones iniciales y que el intervalo de confianza es inversamente proporcional a \sqrt{mn} . Por lo tanto, intervalos mas estrechos se pueden obtener bien sea incrementando m o n . Sin embargo, para reducir el desperdicio (mn_0 observaciones iniciales), se recomienda que m sea pequeño (como 10) y se incremente el valor de n .

2. Medias de tandas

Consiste en correr una simulación prolongada, descartar el intervalo transitorio, y dividir las observaciones restantes en varias tandas o submuestras.

Dada una corrida de $N+n_0$ observaciones, donde n_0 es el tamaño de la fase transitoria y la cual es descartada, se dividen las N observaciones en $m = \lfloor N/n \rfloor$ tandas de n observaciones cada una. Se comienza con valores pequeños de n , por ejemplo $n=1$, y se procede como sigue:

1. Calcule la media de cada tanda:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}, \quad i = 1, 2, \dots, m$$

2. Calcule la media global:

$$\bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

3. Calculemos la varianza de la media de las tandas:

$$Var(\bar{x}) = \frac{1}{m-1} \sum_{i=1}^m (\bar{x}_i - \bar{\bar{x}})^2$$

El intervalo de confianza para la media es

$$\left[\bar{\bar{x}} \pm z_{1-\alpha/2} \sqrt{Var(\bar{x})} \right]$$

Los cálculos son esencialmente los mismos que los anteriores, solo que en este método hay menos desperdicio (solo se descartan n_0 observaciones). El intervalo de confianza también es inversamente proporcional a \sqrt{mn} , y puede ser reducido incrementando m o n . El valor de n debe ser grande para que las medias tengan poca correlación. Una forma de fijar el valor de n consiste en calcular la covarianza de medias de tandas sucesivas:

$$Cov(\bar{x}_i, \bar{x}_{i+1}) = \frac{1}{m-2} \sum_{i=1}^{m-1} (\bar{x}_i - \bar{\bar{x}})(\bar{x}_{i+1} - \bar{\bar{x}})$$

El calculo se repite para valores de n incrementales hasta que la covarianza sea pequeña comparada con la varianza. Se puede comenzar con $n=1$ y sucesivamente ir duplicando el valor de n .

Ejemplo:

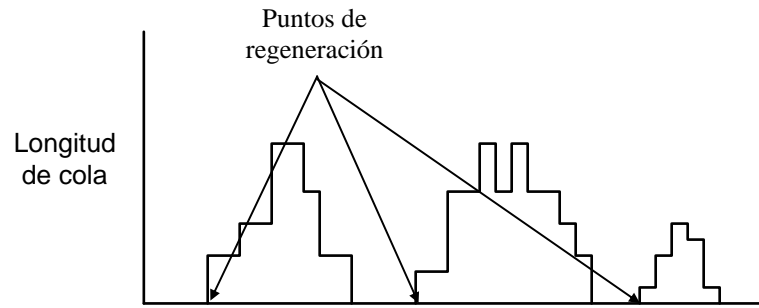
Tamaño de tanda	Covarianza	Varianza	Porcentaje
1	-0.18792	1.79989	10.441
2	0.02643	0.81173	03.256
4	0.11024	0.42003	26.246
8	0.08979	0.26437	33.964
16	0.04001	0.17650	22.669
32	0.01108	0.10833	10.228
64	0.00010	0.06066	00.165
128	-0.00378	0.02992	12.634
256	0.00027	0.01133	02.383
512	0.00069	0.00503	13.718
1024	0.00078	0.00202	38.614

En la tabla se muestran valores de covarianza y varianza para distintos tamaños de tandas. Nótese que para 64 la covarianza es menos del 1% de la varianza. Este es el tamaño de tandas seleccionado.

3. Método de regeneración

Para entender el concepto de regeneración, consideremos un sistema de colas. Comenzando con una longitud de cola cero, la figura muestra una posible trayectoria de la longitud de cola en función del tiempo. Nótese que el sistema frecuentemente retorna al estado inicial. Después de regresar a este estado, la trayectoria no depende de la historia previa. El tiempo de espera depende de las demandas de servicio de trabajos previos, pero si un trabajo llega a un servicio vacío, el tiempo de espera no depende de trabajos anteriores. El intervalo vacío comienza una nueva fase en donde los tiempos de espera no

dependen de algo que sucediera antes del intervalo. Este fenómeno es llamado **regeneración**. Los instantes en los cuales el sistema entra en una fase independiente son llamados **puntos de regeneración**. La duración entre dos puntos de regeneración sucesivos es llamada **ciclo de regeneración**.



La figura muestra tres ciclos. La longitud promedio de la cola del segundo ciclo no esta correlacionada a la del primer ciclo. Este argumento de independendencia también es aplicable a otras variables.

Un **sistema regenerativo** (con ciclos de regeneración) puede ser analizado usando este método. Sin embargo, no todos los sistemas son regenerativos. Un sistema con dos colas se regenera cuando ambas colas se vacían. Al incrementar el numero de colas los puntos de regeneración se vuelven mas escasos y pueden hace el sistema **noregenerativo**.

El cálculo de varianza para este método es un poco más complicado que en los anteriores ya que los ciclos de regeneración son de longitudes diferentes y la media global no puede ser obtenida promediando las medias individuales de los ciclos. Las medias de los ciclos son proporciones con bases (longitudes de ciclo) que son diferentes y el cálculo de la proporción promedio requiere de cuidado.

Supongamos una simulación regenerativa con m ciclos de tamaño n_1, n_2, \dots, n_m , respectivamente. Las medias de los ciclos están dadas por:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

sin embargo la media global no esta dada por la media de la media de los ciclos:

$$\bar{\bar{x}} \neq \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

El procedimiento correcto es:

1. Calculemos la suma de los ciclos:

$$y_i = \sum_{j=1}^{n_i} x_{ij}$$

2. Calculemos la media global:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m n_i}$$

3. Calculemos la diferencia entre sumas de ciclos esperadas y observadas:

$$w_i = y_i - n_i \bar{\bar{x}}, \quad i = 1, 2, \dots, m$$

4. Calculemos la varianza de las diferencias:

$$Var(w) = s_w^2 = \frac{1}{m-1} \sum_{i=1}^m w_i^2$$

5. Calculemos la longitud promedio de los ciclos:

$$\bar{n} = \frac{1}{m} \sum_{i=1}^m n_i$$

El intervalo de confianza para la respuesta media viene dado por:

$$\bar{\bar{x}} \pm z_{1-\alpha/2} \frac{s_w}{\bar{n} \sqrt{m}}$$

Obsérvese que este método no requiere remover el estado transitorio y por lo tanto no hay desperdicio pero tiene un número de desventajas.

1. La longitud de los ciclos es imprevisible y no se puede predecir el tiempo de simulación de antemano.
2. Encontrar los puntos de regeneración no es trivial y puede requerir un montón de chequeos después de cada evento.
3. Muchos de los mecanismos de reducción de varianza como secuencias aleatorias comunes y variables antitéticas no pueden ser usados por la variabilidad de la longitud de los ciclos.
4. Los estimadores de la media y la varianza son sesgados (su valor esperado de una muestra aleatoria no es igual a la cantidad siendo estimada).

Existen otros métodos como el método autoregresivo y el análisis espectral, pero son estadísticamente más sofisticados y no son recomendados para principiantes.

V. REDUCCIÓN DE VARIANZA

Esta técnica requiere controlar secuencias de números aleatorios para introducir correlación entre diferentes corridas para reducir la varianza de los resultados. Reduciendo la varianza se pueden hacer mejores estimaciones sin tener que incrementar el tamaño de la muestra. Un problema con estas técnicas es que su uso inadecuado puede incrementar la varianza. Son para ser usadas por expertos y no por

principiantes.

Hay diferentes técnicas y su aplicación depende del modelo particular y la comprensión que tengamos de el. Entre las técnicas están:

- * Muestreo antitético
- * Secuencias comunes de números aleatorios
- * Variables de control
- * Muestreo estratificado
- * Muestreo selectivo

Discutamos una de ellas.

Muestreo antitético

Se ejecutan pares de simulaciones. Una con números aleatorios r_1 al tiempo t_1 , ..., r_n al tiempo t_n y la segunda simulación corre con una secuencia antitética de números aleatorios $(1 - r_1)$ al tiempo t_1 , ..., $(1 - r_n)$ al tiempo t_n . La intención de la muestra antitética es que produzca una correlación negativa (observaciones pequeñas en una corrida son compensadas por observaciones grandes en la otra) entre las variables de las dos simulaciones.

Supongamos que repetimos N veces una simulación. Tendremos que:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{y} \quad \text{Var}(\bar{x}) = \frac{1}{N} \left[\sum_{i=1}^N \text{var}(x_i) + 2 \sum_{i=1}^N \sum_{j=i+1}^N \text{Cov}(x_i, x_j) \right]$$

Si las corridas son independientes, la covarianza es cero; pero si la suma de las covarianzas es negativa, entonces la varianza de \bar{x} se reduce.