

## Diseño del experimento de simulación y análisis de resultados. (Thesen & Travis)<sup>1</sup>

Una de las razones por las cuales llevar a cabo un *estudio de simulación* es para determinar el *desempeño* (“performance”) de un sistema para un conjunto de condiciones dadas. El analista de simulación debe tomar una serie de *decisiones* acerca de cómo será recogida y analizada la *información proveniente de la simulación*.

En la fase de recolección de información, el analista debe decidir, ¿bajo qué condiciones simular el sistema?, ¿cuántas réplicas llevar a cabo?, ¿qué duración tendrá cada réplica?. Lo anterior y otras preguntas relacionadas se refiere al “**diseño del experimento de simulación**”.

Una vez diseñado e implementado el experimento, el analista extrae la información para la *toma de decisiones* a partir de la información resultante de la simulación, usando técnicas estadísticas comúnmente denominadas “**análisis de resultados**”.

El “*diseño del experimento de simulación*” y el “*análisis de resultados*” están relacionados entre sí, ya que el primero debe adaptarse al segundo. En otras palabras el experimento de simulación debe ser diseñado para producir la información apropiada al tipo de análisis (anticipado) de resultados. Debemos distinguir entre dos tipos de simulación:

1. *Simulación de terminación*. Aquella que se ejecuta durante un tiempo  $T_E$  donde  $E$  es un evento(s) que detiene(n) la simulación.  $E$  puede ser un tiempo especificado o una condición especificada. *Ejemplo*, un sistema de líneas de espera que se detiene después de haber observado  $N$  clientes.
2. *Simulación de estado estable*. Aquella que se ejecuta durante un tiempo largo (la duración “tiende a infinito”). Se evalúa el comportamiento del sistema cuando este ha sido operado bajo condiciones fijas por un tiempo relativamente largo. *Ejemplo*, evaluar el desempeño en estado estable de una sala de emergencias de un hospital.

### ***Simulación de terminación.***

La medida general de la variabilidad se enuncia a menudo en la forma de un **intervalo de confianza** a determinado **nivel de confianza**. Se llama intervalo de confianza para el parámetro  $\theta$  con nivel de confianza  $(1 - \alpha)$ , a la expresión<sup>2</sup>,  $\theta_I \leq \theta \leq \theta_S$ , es decir,

$$\begin{aligned} P(\theta_I \leq \theta \leq \theta_S) &= 1 - \alpha \quad (\text{intervalo de confianza bilateral}) \\ \left. \begin{aligned} P(\theta_I < \theta) &= 1 - \alpha \\ P(\theta < \theta_S) &= 1 - \alpha \end{aligned} \right\} & (\text{intervalo de confianza unilateral}) \end{aligned}$$

<sup>1</sup> Arne Thesen, Laurel E. Travis, “Simulation for Decision Making”, West Publishing Company, 1992.

<sup>2</sup>  $I$  = inferior.  $S$  = superior.

Por lo general se desea estimar el “valor real” de la media  $\mu$  (valor promedio) de la medida de desempeño en cuestión (esto es imposible). Lo mejor que podemos hacer es calcular un intervalo que incluya tal valor. Se suele usar  $\bar{x}$  como estimador de  $\mu$  y calcularemos una cierta medida de amplitud  $\delta$  del intervalo (distancia a ambos lados de  $\bar{x}$ ) para un determinado “nivel de riesgo”  $\alpha$  (llamado también *nivel de significación* o *nivel de confianza*).

$$P(\bar{x} - \delta < \mu < \bar{x} + \delta) = 1 - \alpha$$

*Ejemplo:*

Se tiene un sistema sencillo de líneas de espera (una cola y un servidor) donde nos interesa estimar el *tiempo de espera promedio* (medida de desempeño) de los primeros  $N$  ( $=10$ ) clientes. Como cada ejecución del modelo de simulación nos puede dar diferentes valores de la medida de desempeño, nos interesa estimar el “valor real” (valor esperado) de la medida de desempeño y el rango en el que se podría encontrar tal “valor real”.

Se han obtenido 200 valores y como las réplicas (ejecución del modelo de simulación) son independientes, la estadística<sup>3</sup> nos provee de la siguiente expresión para el cálculo de  $\delta$ .

$$\delta = t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}}$$

donde,

$n$  es el número de observaciones.

$$s \text{ es la desviación estándar de la muestra (estimador insesgado)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}.$$

$\bar{x}$  es el valor promedio de la muestra.

$gl$  son los grados de libertad<sup>4</sup> ( $gl = n - 1$ ).

$\alpha$  es el nivel de significación, nivel de confianza o nivel de riesgo.

$t_{1-\alpha/2, gl}$  es el valor apropiado de la distribución  $t$  de student.

Cuando el número de grados de libertad es grande ( $gl \rightarrow \infty$ ), la distribución  $t$  de student se aproxima a una distribución normal y en consecuencia, se utiliza el valor correspondiente de la normal  $z_{1-\alpha/2}$  en lugar de  $t_{1-\alpha/2, gl}$ . Para fines prácticos cuando  $gl \geq 30$ , la aproximación es aceptable<sup>5</sup>.

<sup>3</sup> Véase la NOTA 1 (al final del documento).

<sup>4</sup> Véase la NOTA 2 (al final del documento).

<sup>5</sup> J. Prawda W., “Métodos y Modelos de Investigación de Operaciones”, Vol. II. Editorial Limusa, 1991.

En el presente ejemplo tenemos que,  $n = 200$ ,  $\bar{x} = 2.07$  y  $s = 1.21$ . Si estamos dispuestos a aceptar que nuestro intervalo no incluya el valor de la media con un riesgo del 10%, esto quiere decir que calcularemos  $\delta$  para un intervalo de confianza al 90%.

$$\delta = t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}} = t_{0.95, 199} \frac{1.21}{\sqrt{200}} = 1.645 \frac{1.21}{\sqrt{200}} = 0.14$$

Luego, el intervalo de confianza al 90% será,

$$(\bar{x} - \delta, \bar{x} + \delta) = (2.07 - 0.14, 2.07 + 0.14) = (1.93, 2.21).$$

En general, se pueden estimar intervalos de confianza para una simulación de terminación de la siguiente manera:

- 1: Establecer un nivel aceptable de riesgo, por ejemplo,  $\alpha = 10\%$ .
- 2: Recoger una muestra de  $n$  observaciones  $(x_i, i = 1, 2, \dots, n)$  que satisfaga dos condiciones:  
Independencia: Se garantiza la independencia, en las simulaciones de terminación donde se recoge una observación para cada réplica y el sistema es reinicializado a partir de un estado específico. Todas las variables estado son inicializadas nuevamente entre réplicas sucesivas.  
Normalidad: Si cada observación individual es un promedio de un gran número ( $n \geq 30$ ) de observaciones independientes provenientes de una misma distribución, el Teorema del Límite Central<sup>6</sup> nos dice que el promedio de esas observaciones tendrá una distribución aproximadamente normal.
- 3: Calcule la media  $(\bar{x})$  y la varianza  $(s^2)$  de la muestra.
- 4: Establezca  $gl = n - 1$  y calcule  $\delta$ .  

$$\delta = t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}}$$
- 5: Calcule los límites inferior y superior del intervalo de confianza.  

$$x_I = \bar{x} - \delta$$

$$x_S = \bar{x} + \delta$$
- 6: Concluya que existe una probabilidad  $(1 - \alpha)$  que el intervalo de confianza  $(\bar{x} - \delta, \bar{x} + \delta)$  incluya el valor real de la media.

Observe que el  $\delta$  ("**precisión**") del intervalo de confianza resultante no se puede conocer a priori, es decir, hasta que todos los cálculos correspondientes sean realizados. Sin

<sup>6</sup> Véase la NOTA 3 (al final del documento).

embargo, para efectos de la toma de decisiones seria importante poder definir a priori la *precisión* de nuestra estimación.

### Intervalos de confianza para un $\delta$ dado.

Al momento de determinar el **número de réplicas** que deben realizarse para una **simulación de terminación dada**, se deben tener en cuenta dos consideraciones:

- El costo de cada réplica.
- La precisión ( $\delta$ ) del intervalo de confianza resultante.

Supongamos que nos interesa encontrar un intervalo de confianza al 95% alrededor de la media  $\mu$  con una precisión no mayor a 7 unidades, ¿Cuántas réplicas serán necesarias?. El problema que deseamos resolver consiste en despejar  $n$  de la siguiente ecuación

$$\delta = t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}}.$$

Sin embargo, no es tan sencillo pues  $s$  es función de la muestra (aleatoria) y en consecuencia es una variable aleatoria. Se necesitan datos para poder calcular  $s$ . Para resolver este problema (estimación de  $s$ ), podemos recoger una muestra inicial para una estimación inicial de  $s$  previo al cálculo del tamaño “deseado” de la muestra.

Otro problema adicional es el valor  $t_{1-\alpha/2, gl}$  pues se necesita saber el número de grados de libertad ( $gl$ ), que es función de  $n$  también. Recordemos que cuando el tamaño de la muestra es grande,  $t_{1-\alpha/2, gl}$  puede ser aproximado por el valor equivalente correspondiente a la distribución normal  $z_{1-\alpha/2}$ . En consecuencia, si obtenemos una muestra inicial con una desviación estándar de la muestra  $s_0$ , podemos resolver la ecuación para obtener un valor aproximado para  $n$ .

$$\delta = z_{1-\alpha/2} \frac{s_0}{\sqrt{n}} \Rightarrow n = \left( z_{1-\alpha/2} \frac{s_0}{\delta} \right)^2$$

Luego, dado un cierto nivel de confianza  $\alpha$  y un  $\delta$  deseado ( $\delta_{\text{objetivo}}$ ), podemos determinar el número de réplicas (número de ejecuciones de la simulación) mediante el siguiente procedimiento:

**1:** Seleccione el número de réplicas a realizar de acuerdo con una de las siguientes reglas y proceda.

- Si se trata de la **primera** vez que se recogen datos, se debe realizar un número pequeño de réplicas (entre 5 y 30 dependiendo del “costo”).
- Si se trata de la **segunda** vez que se recogen datos, realice suficientes réplicas para alcanzar el total  $n/2$ , donde  $n$  ha sido calculado la última vez que el paso 3 haya sido realizado.
- Si se han recogido datos en **más de dos** oportunidades, realice suficientes réplicas

para alcanzar el total  $n$ , donde  $n$  ha sido calculado la última vez que el paso 3 haya sido realizado.

- 2:** Calcule la media de la muestra  $\bar{x}$  y la desviación  $s$  a partir del conjunto de datos obtenidos en el paso 1. A continuación calcule  $\delta$  mediante la fórmula

$$\delta = t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}.$$

- 3:** Si el valor de  $\delta$  calculado es menor o igual que  $\delta_{\text{objetivo}}$  entonces el procedimiento habrá concluido. En caso contrario, se necesitarán más datos para alcanzar la precisión deseada. Utilice el valor actual de  $s$  y calcule nuevamente el número de réplicas  $n$  mediante la fórmula

$$n = \left( z_{1-\alpha/2} \frac{s}{\delta_{\text{objetivo}}} \right)^2$$

y regrese al paso 1.

#### Ejemplo:

Se desea estimar el *tiempo de espera promedio* para los primeros  $N (=10)$  clientes en un sistema de líneas de espera con una cola y un servidor. Los tiempos entre llegadas y los tiempos de servicio se distribuyen exponencial. Se trata de una simulación de terminación. Se desea obtener un intervalo de confianza al 10% ( $\alpha = 0.10$ ) con un  $\delta$  no mayor de 0.15.

- 1:** Recolectamos datos para un número pequeño de réplicas ( $n = 20$ ). En cada réplica, se calcula el *tiempo de espera promedio* para los primeros 10 clientes. Se obtuvieron las siguientes observaciones.

1,71	1,69	1,17	1,35	0,71	0,72	0,56	1,4	1,37	0,61
1,24	0,81	2,57	1	1,49	1,41	1,27	3,58	0,51	2,03

- 2:** Calculamos el estimado inicial de  $\bar{x}$ ,  $s$  y  $\delta$ .

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 1.36$$

$$s_0 = \sqrt{\frac{1}{19} \sum_{i=1}^{20} (x_i - \bar{x})^2} = 0.738$$

$$\delta = t_{0.95, 19} \frac{s}{\sqrt{n}} = 1.729 \frac{0.738}{\sqrt{20}} = 0.285$$

- 3:** Aún no hemos alcanzado el nivel de precisión deseado ya que el  $\delta$  calculado en 2 es mayor 0.15 (la muestra no es lo suficientemente grande). Estimamos el número total de datos necesarios calculando.

$$n = \left( z_{1-\alpha/2} \frac{s_0}{\delta_{\text{objetivo}}} \right)^2 = \left( 1.645 \frac{0.738}{0.15} \right)^2 = 65.5$$

Redondeamos la aproximación obtenida (66) en el paso 3 y repetimos el paso 1 del procedimiento anterior.

**1:** Realizamos ejecuciones adicionales (nuevas réplicas) hasta alcanzar el total 66/2, es decir,  $(66/2) - 20 = 13$  réplicas adicionales.

**2:** Calculamos nuevamente  $\bar{x}$ ,  $s$  y  $\delta$ .

$$\begin{aligned}\bar{x} &= 1.51 \\ s &= 1.04 \\ \delta &= t_{0.95,32} \frac{s}{\sqrt{n}} = 1.69 \frac{1.04}{\sqrt{33}} = 0.31\end{aligned}$$

**3:** Como todavía 0.31 es mayor que el  $\delta_{\text{objetivo}}$  de 0.15, estimamos nuevamente  $n$

$$n = \left( z_{1-\alpha/2} \frac{s}{\delta_{\text{objetivo}}} \right)^2 = \left( 1.645 \frac{1.04}{0.15} \right)^2 = 130$$

y regresamos al paso 1.

**1:** Se ejecutan  $130 - 33 = 97$  ejecuciones adicionales del modelo de simulación.

**2:** Se calculan nuevamente  $\bar{x}$ ,  $s$  y  $\delta$  para nuestro conjunto de 130 datos.

$$\begin{aligned}\bar{x} &= 1.42 \\ s &= 0.84 \\ \delta &= 1.66 \frac{0.84}{\sqrt{130}} = 0.12\end{aligned}$$

**3:** Terminamos el análisis en este punto pues el intervalo de confianza cumple con las condiciones solicitadas.

Se concluye que existe una probabilidad del 90% que el intervalo  $(1.42 - 0.12, 1.42 + 0.12)$  incluya el valor real de la media de nuestra medida de desempeño. Se necesitan 130 réplicas para obtener suficiente información y calcular un intervalo de confianza del tamaño deseado.

### ***Simulación de estado estable.***

El análisis de datos en el caso de las simulaciones de estado estable presenta dos dificultades.

- *Primero*, debemos poder asegurar que el sistema se encuentra en estado estable antes de empezar a recolectar datos. En caso de no poder hacerlo, los datos presentarán un problema conocido como **sesgo inicial**.
- *Segundo*, antes de realizar el cálculo de los intervalos de confianza, debemos estar seguros que nuestros datos satisfacen la hipótesis de **independencia**.

### Sesgo inicial.

Fallos en la consideración del sesgo inicial al momento de diseñar el experimento de simulación pueden conducir a una subestimación o una sobrestimación de la medida de desempeño en cuestión. En teoría, el sesgo inicial no puede ser eliminado por completo. Sin embargo, en la práctica se puede reducir su impacto a través de:

- La introducción de periodos de calentamiento.
- Empezando la simulación en un estado más verosímil.

En muchas situaciones la experiencia práctica de algunos sistemas puede servir como *guía* a la hora de determinar periodos de calentamiento o el estado inicial del sistema. Durante el periodo de calentamiento no se recolectan datos. La recolección de datos empieza una vez que el periodo de calentamiento ha terminado. Se puede emplear el método de **réplicas independientes** o el método de **una larga ejecución (corrida)**.

#### 1: *Réplicas independientes.*

Se ejecutan simulaciones recolectando una observación (un dato, una medida de desempeño) por simulación para después utilizar el conjunto de datos y calcular el intervalo de confianza correspondiente.

##### **Problemas.**

- ¿Cuál debe ser la longitud de los periodos de calentamiento?.
- ¿El periodo de calentamiento especificado podría no ser lo suficientemente largo para eliminar el sesgo inicial?.
- El uso repetido de largos periodos de calentamiento implica una gran pérdida de datos.

#### 2: *Una larga corrida (batch means technique).*

Todas las observaciones se realizan en una ejecución o corrida simple. Después del periodo de calentamiento, se recolectan los datos a intervalos especificados (batches).

##### **Problema.**

Al no reiniciar el sistema cada vez que se recogen datos, no podemos garantizar que las observaciones sean independientes.

### Independencia.

Siempre debemos asegurarnos que los datos recolectados cumplan con la hipótesis de independencia antes de calcular intervalos de confianza. En el caso de las simulaciones de terminación y de las simulaciones de estado estable usando el método de las réplicas independientes no hay problema para garantizar la independencia. Sin embargo, en el método de una larga corrida (batch means technique) la independencia es una dificultad. La

dependencia entre datos sucesivos puede ser reducida usando un tamaño de “batch” más grande<sup>7</sup>.

### NOTA 1

Si  $X$  tiene una distribución normal con  $\mu$  y  $\sigma$ , entonces  $Z = \frac{x - \mu}{\sigma}$  tiene una distribución Normal(0,1). Por lo general, no se conoce la  $\sigma$  de la población y se tiene que inferir a través de  $s$  (la desviación estándar de la muestra). La estadística  $t = \frac{(\bar{x} - \mu)\sqrt{n}}{s}$  tiene una distribución  $t$  de student con  $gl = n-1$ .

$$P\left(-t_{1-\alpha/2, gl} < \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} < t_{1-\alpha/2, gl}\right) = 1 - \alpha$$

$$P\left(\bar{x} - t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2, gl} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

$$P(\bar{x} - \delta < \mu < \bar{x} + \delta) = 1 - \alpha$$

### NOTA 2

La noción (idea) de “grado de libertad” tiene en cuenta el número de términos desconocidos antes de tomar la muestra que incluimos en el cálculo del estimador, por

ejemplo, la varianza de la muestra,  $s^2 = \frac{\sum_{i=1}^n e_i^2}{n-1}$ , donde  $e_i = x_i - \bar{x}$  (error).

$$n = 1, \bar{x} = x_1 \quad e_1 = 0$$

$$n = 2, e_1 = x_1 - \bar{x} = x_1 - \left(\frac{x_1 + x_2}{2}\right) = \frac{x_1 - x_2}{2},$$

$$e_2 = x_2 - \bar{x} = x_2 - \left(\frac{x_1 + x_2}{2}\right) = \frac{x_2 - x_1}{2} = -\frac{(x_1 - x_2)}{2} = -e_1$$

Un solo grado de libertad,  $e_1$  ó  $e_2$ . Dado alguno de ellos, el otro queda automáticamente determinado.

$$n = 3, e_1 = x_1 - \bar{x} = x_1 - \left(\frac{x_1 + x_2 + x_3}{3}\right) = \frac{2x_1 - x_2 - x_3}{3},$$

<sup>7</sup> Arne Thesen, Laurel E. Travis, “Simulation for Decision Making”, West Publishing Company, 1992 (pp 169-197).



$$e_2 = x_2 - \bar{x} = x_2 - \left( \frac{x_1 + x_2 + x_3}{3} \right) = \frac{-x_1 + 2x_2 - x_3}{3},$$

$$e_1 + e_2 = \frac{x_1 + x_2 - 2x_3}{3} = -e_3$$

Dos grados de libertad. En general,  $\sum (x_i - \bar{x}) = \sum e_i = 0$ , solo hay  $(n-1)$  residuos desconocidos pues el último puede calcularse a partir de los otros.

### NOTA 3<sup>8</sup>

Si se tiene una muestra aleatoria  $(x_1, x_2, \dots, x_n)$  de tamaño  $n$  de una distribución normal con media  $\mu$  y varianza  $\sigma^2$ , entonces  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  tiene una distribución normal con media  $\mu$  y varianza  $\sigma^2/n$ .

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu$$

$$Var(\bar{x}) = Var\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(x_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Si  $\bar{x}$  presenta una distribución normal con media  $\mu$  y varianza  $\sigma^2/n$ , entonces  $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \left( \frac{\bar{x} - \mu}{\sigma} \right)$  tiene una distribución normal estándar, es decir,  $N(0,1)$ .

Lo anterior nos demuestra que si se toma una muestra de una población normal entonces  $\bar{x}$  tiene una distribución muestral normal. ¿Qué podemos decir de la distribución muestral de  $\bar{x}$  si los  $x_i$  no se distribuyen normalmente?. Para dar respuesta a lo anterior se tiene el Teorema del Límite Central:

Sean  $x_1, x_2, \dots, x_n$  variables aleatorias independientes y distribuidas idénticamente con  $E(x_i) = \mu$  y  $Var(x_i) = \sigma^2 < \infty$ . Definimos

$$U_n = \sqrt{n} \left( \frac{\bar{x} - \mu}{\sigma} \right) \text{ en donde } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Entonces la función de distribución  $U_n$  converge a una función de distribución normal estándar cuando  $n \rightarrow \infty$ .

“ $\bar{x}$  tendrá una distribución muestral aproximadamente normal si el tamaño de la muestra es grande ( $n \geq 30$ )”.

<sup>8</sup> W. Mendenhall, Richard L. Scheaffer, Dennis D. Wackerly, “Estadística Matemática con Aplicaciones”, Grupo Editorial Iberoamérica, México, 1986.