PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Escuela de Ingeniería
Integrante: Joaquin Ossandón Stanke.
Profesor: Pedro Gazmuri.
Ramo: Fundamentos de simulación - ICS2133.

Tarea 3

1.

Desarrolle un modelo en SIMIO para el problema de la Tarea N1 (supermercado). Realice 20 réplicas. Calcule el intervalo de confianza de la medida de desempeño en base a los dos intervalos propuestos en clases. Calcule el error de la estimación en cada caso.

A continuación desarrollaré una breve descripción de mi programa en SIMIO, para así tener una mejor comprensión a la hora de evaluar en el mismo programa. En primer lugar, creé el *server* de llegada, donde clientes ingresaban a una tasa exponencial con media 1/3. Después de esto, pasaban a otro *server* donde demoraban un proceso aleatorio exponencial de media 15 minutos. Después de transcurrido que alguno de los clientes comprara, estos pasaban a un *Transfer Node*, el cual seleccionaba de una lista, que tenía las 16 cajas del Supermercado, a la cola más corta tomando el *Smallest Value* de los valores de la lista, obtenidos mediante el comando **Candidate.Server.InputBuffer.Contents.NumberWaiting + Candidate.Server.Processing.Contents.NumberWaiting**, el cual suma la cantidad de personas esperando en la cola y la cantidad de personas esperando en el proceso de pago en la caja (que tiene como límite 1). De esta forma, el cliente decide a dónde ir, dependiendo de la cantidad de gente que hay en cada cola, tomando como preferencia desde la Caja 1 hasta la Caja 16, en orden ascendente. Finalmente después de un proceso aleatorio exponencial de media 5 minutos, se procede a salir del Supermercado.
Es importante notar que, si bien la Supermercado cierra a las 14 horas de jornada, la simulación dura 18 horas, y la tasa de llegada de personas se corta a las 14 horas, simulando que existen personas que quedan dentro del Supermercado. La figura a continuación muestra el programa simulando.
Nota: el modelo entregado en SIDING no tiene los *Symbols* agregados en la figura siguiente, ya que es puramente funcional (se hizo en la Universidad); el de la imagen se realizó y no se comprobó el funcionamiento óptimo por tema de licencia.

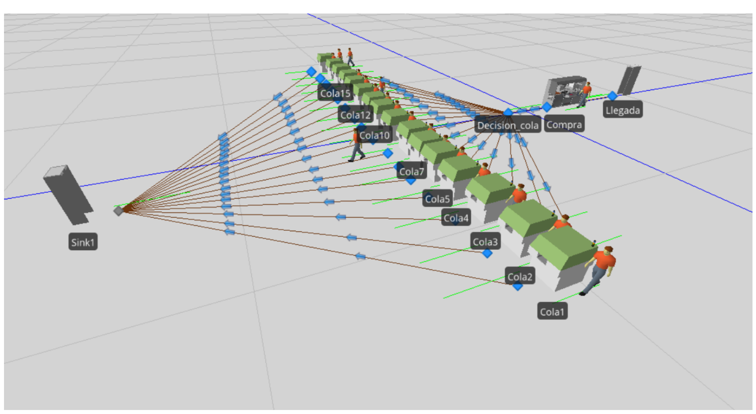

Figura 1: Ilustración de la simulación del supermercado en *SIMIO*.

Desarrollando las 20 réplicas del modelo, se obtuvo un promedio de 4.66 minutos de espera en la cola. El modelo se puede observar en detalle en el archivo *.spfx*. A continuación se muestran los tiempos promedios en la cola para cada caja. Además, se muestran los intervalos de confianza de la medida de desempeño en base a los dos intervalos propuestos en clases para μ y σ desconocidos.

Cola	Tiempo promedio (*mins.*)
1	6.444
2	6.335
3	5.929
4	5.692
5	5.2615
6	5.124
7	4.3179
8	4.2721
9	4.1659
10	3.959
11	3.874
12	3.869
13	3.9032
14	3.837
15	3.967
16	3.692
Mean	4.6651

Tabla 1: Resultados de la simulación en SIMIO.

Método 1:

Sea $X_1, X_2, ..., X_n$ v.a. iid cuando $\frac{\overline{X}(n)-\mu}{\sqrt{\frac{S^2(n)}{n}}} \xrightarrow{n \longrightarrow \infty} N(0,1)$. De esta forma el intervalo de confianza para $1-\alpha = 0{,}05$ queda como sigue.

$$\overline{X}(n) - 1{,}96 \cdot \sqrt{\frac{S^2(n)}{n}} < \mu < \overline{X}(n) + 1{,}96 \cdot \sqrt{\frac{S^2(n)}{n}}$$

$$4{,}22815 < \mu < 5{,}102128$$

Presentando un error de estimación de 0,094.

Método 2:

Si asumimos que $X_1, X_2, ..., X_n$ v.a. iid con $X_i \tilde{} N(\mu, \sigma^2)$. De esta forma el intervalo de confianza para $1-\alpha = 0{,}05$ queda como sigue.

$$\overline{X}(n) - 2{,}093 \cdot \sqrt{\frac{S^2(n)}{n}} < \mu < \overline{X}(n) + 2{,}093 \cdot \sqrt{\frac{S^2(n)}{n}}$$

$$4{,}1984 < \mu < 5{,}1317$$

Presentando un error de estimación de 0,1.

Donde 1,96 y 2,093 son $Z_{1-\alpha}$ y $T - Student_{19grados}$ respectivamente. Además de tener que $S = 0{,}9718$ por los datos obtenidos (los promedios de espera por cada cola).

2.

Considere una v.a. uniforme entre a y b. Discuta métodos para estimar a y b en base a una muestra aleatoria.

El primer método a utilizar será el estimador de momentos.
Tenemos que, dados los momentos de una Uniforme(a, b):

- Primer momento: $\frac{a+b}{2} = E(X)$
- Segundo momento: $\frac{a^2+ba+b^2}{3} = E(X^2)$

Dado esto hacemos

- $\overline{X} = E(X) = \frac{a+b}{2}$
- $\text{Var}(X_1...X_n) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}$

Despejando el sistema de ecuaciones para a, b y considerando la fórmula de varianza vista en clases, obtenemos finalmente

- $\hat{a} = \overline{X} - \sqrt{3(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2)}$
- $\hat{b} = \overline{X} + \sqrt{3(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2)}$

Si bien este método es más accesible y fácil calcular a mano, en algunos casos cuando estimamos parámetros de una distribución de probabilidad conocida, como lo es de una Uniforme, es mejor sustituir este método por Máxima Verosimilitud, ya que así existe una mayor probabilidad de que los valores sean más cercanos a las cantidades que estimamos y sean menos sesgadas. Ahora veamos el método de Máxima Verosimilitud (EMV).
Sea $X_1...X_n$ - $U(a, b)$

$$p(X|a, b) = \frac{1}{b-a} \text{ con } a < X < b$$

Dado esto, definimos lo que debemos maximizar

$$L(b, a|X_1, ..., X_n) = \frac{1}{(b-a)^n}$$

Notemos que L es monótona decreciente en función de b. Luego L se incrementa al disminuir b, por lo que b debe ser el menor valor posible, y por tanto, la mayor observacion del *set* $X_1, ..., X_n$ (ya que es el menor valor posible que podemos tomar de los conocidos). Esto ocurre de igual forma para a, L se incrementa al aumentar a, por lo que debemos aumentarlo de tal forma que no sobrepase cualquiera de las observaciones, de esta forma será la menor observación del *set* $X_1, ..., X_n$.

Así, nos queda que

- $\hat{a} = min\{X_1, \ldots, X_n\}$
- $\hat{b} = max\{X_1, \ldots, X_n\}$

Este método es conocido por su insesgamiento cuando n (la cantidad de observaciones) es grande. Bajo estas condiciones, el estimador es consistente y eficiente.

Como tercer y último método, se propone obtener las distribuciones del máximo y mínimo de las muestras aleatorias, para luego estimar los parametros a y b en función de la esperanza de estas distribuciones y los valores límite obtenidos a partir de la muestra. Dado esto, procedemos a calcular las distribuciones.

Sea $Z = max\{X_1, ..., X_n\}$. Entonces la distribucion de Z está dada por

$$P(Z \leq x) = P(X_1 \leq x, ..., X_n \leq x)$$

$$P(Z \leq x) = P(X_1 \leq x) \cdot ... \cdot P(X_n \leq x)$$

$$P(Z \leq x) = (\frac{x-a}{b-a})^n$$

Dado esto, tenemos que

- $F_z(x) = (\frac{x-a}{b-a})^n$
- $f_z(x) = n(\frac{x-a}{b-a})^{n-1} \cdot \frac{1}{b-a}$

Ahora veamos lo mismo, pero con el mínimo. Sea $W = min\{X_1, ..., X_n\}$. Entonces la distribucion de W está dada por

$$P(W \geq x) = P(X_1 \geq x, ..., X_n \geq x)$$

$$P(W \geq x) = P(X_1 \geq x) \cdot ... \cdot P(X_n \geq x)$$

$$P(W \geq x) = (1 - \frac{x-a}{b-a})^n$$

$$P(W \leq x) = 1 - (1 - \frac{x-a}{b-a})^n$$

Dado esto, tenemos que

- $F_w(x) = 1 - (\frac{b-x}{b-a})^n$
- $f_w(x) = n(\frac{b-x}{b-a})^{n-1} \cdot \frac{1}{b-a}$

Teniendo ambas distibuciones, procedemos a calcular los valores esperados, que luego reemplazaremos por los valores máximo y mínimo de la muestra. De esta forma, tendremos un sistema de ecuaciones para a y b, dado una muestra con n observaciones.

Así, calculamos las integrales respectivas (la integral de x por la distribución). A continuación se muestran las integrales a desarrollar reescritas junto con sus resultados.

- $E_(x_{max}) = \int_a^b \frac{nx\left(\frac{x-a}{b-a}\right)^{n-1}}{b-a} = \frac{(bn+a)}{(n+1)}$
- $E_(x_{min}) = \int_a^b \frac{n\left(\frac{b-x}{b-a}\right)^{n-1}x}{b-a} = \frac{(an+b)}{(n+1)}$

Resolviendo y despejando términos, obtenemos

- $\hat{a} = \frac{nE(x_{min}) - E(x_{max})}{n-1}$
- $\hat{b} = \frac{nE(x_{max}) - E(x_{min})}{n-1}$

Teniendo esto, tomamos $E_(x_{max}), E_(x_{min})$ como los valores máximos y mínimos de la muestra, respectivamente, y resolvemos el sistema de ecuaciones. Notemos que, si hacemos n tender a a infinito en ambos resultados, se tendra que la esperanza del máximo de las variables será b, mientras que la esperanza del mínimo de las variables será a, teniendo así que este tiene los mismos resultados que MV con n grande.

Ahora, tomemos un pequeño ejemplo en donde podemos diferenciar MV de este último método. Supongamos el escenario a continuación, donde tenemos distintas observaciones de una distribución Uniforme entre 5 y 30.

	Uniforme entre 5 y 30			
22,095522	26,286902	N	Máximo	Mínimo
14,7031341	19,0194739	12	26,286902	6,76808409
12,2748965	16,8907925		*a* estimado	*b* estimado
15,3767014	11,9181532	MV	6,76808409	26,286902
25,7626214	6,76808409	MAX - MIN	4,9936461	28,06134
12,0455948	25,3256665			

Figura 2: Instancias Uniformes entre 5 y 30 con máximo y mínimo de la muestra.

Si utilizaramos el método de Máxima Verosimilitud (MV) en este caso, tomaríamos $a = 6{,}76$ y $b = 26{,}28$, dadas las condiciones que impuse más atras. Por otro lado, si usamos las estimaciones con el último método, y reemplazamos $E(x_{min}), E(x_{max})$ por el mínimo y máximo de la tabla respectivamente, obtenemos que $\hat{a} = 4{,}99$ y $\hat{b} = 28{,}06$, algo mucho más cercano a 5 y 30.
Es importante notar que esto no ocurrirá siempre, ya que depende de las observaciones. Si casualmente el máximo y el mínimo de la muestra son parecidos a los límites teóricos, entonces MV sobrepasará al último método notoriamente, no al contrario (como lo acabamos de ver). Es por esto que **asintóticamente** MV será más eficiente, ya que existirán más datos cercanos a los parámetros teóricos.

3.

La Tabla adjunta contiene una muestra de 50 valores de tiempos ente entradas a un baño en un centro comercial (tiempos en segundos). Se quiere aplicar el test de Chi cuadrado para verificar si la distribución asociada a estos tiempos es la exponencial. Defina 15 intervalos para hace el test. Defina en detalle el procedimiento para especificar los intervalos y para calcular el estadígrafo, y para aplicar el test.

En primer lugar calculamos el parámetro de la distribución exponencial λ con los 50 valores que tenemos. Esto lo hacemos de la forma $\frac{1}{Mean}$ con $Mean$ la media de las observaciones entregadas otorgándonos un valor de 0,1022.
Luego, como tenemos 15 intervalos, calculamos $1/15 = 0{,}0666$ que ajustamos como probabilidad de caer en alguno de los intervalos que crearemos. Dado esto, procedemos a crear los intervalos.

Notemos que la probabilidad de que una exponencial esté en el primer intervalo está dado por

$$F(X_1) - F(X_0) = 0{,}0666$$

Sabemos que una exponencial parte del 0, por lo que esta expresión queda como

$$1 - e^{-0{,}1022 \cdot X_1} = 0{,}0666$$

$$X_1 = 0{,}6749$$

Siguiendo con la secuencia, obtenemos lo siguiente

Intervalo 1	0	0.6749
Intervalo 2	0.6749	1.399
Intervalo 3	1.399	2.1828
Intervalo 4	2.1828	3.034
Intervalo 5	3.034	3.966
Intervalo 6	3.966	4.997
Intervalo 7	4.997	6.1493
Intervalo 8	6.1493	7.455
Intervalo 9	7.455	8.9635
Intervalo 10	8.9635	10.747
Intervalo 11	10.747	12.929
Intervalo 12	12.929	15.744
Intervalo 13	15.744	19.710
Intervalo 14	19.710	26.491
Intervalo 15	26.491	50

Tabla 2: Intervalos creados para la muestra

Donde se puede apreciar que se puso como límite 50 dado que abarca todos los valores entregados. Ahora procederemos a calcular el estadígrafo. Para esto definimos:

- N_j: Cantidad de valores de la muestra que están en el intervalo j.
- P_j: Probabilidad de que algún valor este en el intervalo j.

De esta forma usamos la fórmula

$$\chi^2 = \frac{(N_j - nP_j)^2}{nP_j}$$

Nótese que se usó como $P_j = 1/15\ \forall j$, y los N_j fueron calculados con *Excel* y la función **COUNTIF** para las observaciones dadas.

Reemplazando en el estadígrafo, entrega como resultado 16,6. Procedemos a comparar con el valor en tabla para $(15 - 1 - 1)$ grados de libertad (ya que estimamos un parámetro) para una significancia del 95 %, obteniendo que el número a comparar es 22.362, por lo que **no** se rechaza la hipótesis nula de que las distribuciones son iguales (ya que $\chi^2 < \chi^2_{95\,\%}(13)$).

Referencias

[1] Castro-Kuriss, Claudia, *Distribución del Máximo y del Mínimo de una muestra aleatoria de una variable aleatoria X*, Buenos Aires, Argentina, 2017.

[2] Statistics I: Course Material, *Point estimation*, University of London, 2003.