Práctico 6 - Análisis estadístico de datos simulados

Ejercicio 1. Genere n valores de una variable aleatoria normal estándar de manera tal que se cumplan las condiciones: $n \geq 100$ y $S/\sqrt{n} < 0{,}1$, siendo S el estimador de la desviación estándar de los n datos generados.

a) ¿Cuál es el número de datos generados efectivamente?

b) ¿Cuál es la media muestral de los datos generados?

c) ¿Cuál es la varianza muestral de los datos generados?

Ejercicio 2. Estime mediante el método de Monte Carlo la integral

$$\textbf{i)} \int_0^1 \frac{e^x}{\sqrt{2x}}\,dx, \qquad \textbf{ii)} \int_{-\infty}^{\infty} x^2 \exp(-x^2)\,dx.$$

a) Indique cómo se obtiene mediante simulación el valor de la integral.

b) Genere al menos 100 valores y deténgase cuando la desviación estándar muestral S del estimador sea menor que 0,01.

a) En el caso de la integral ii primero se debe realizar una sustitución para obtener una integral de la forma $\int_0^1 g(x)\,dx$

Luego en ambos casos generamos N variables aleatorias uniformes $\mathcal{U}(0,1)$ y calculamos la media muestral de evaluar cada variable en $g(x)$, es decir

$$\int_0^1 g(x)\,dx \simeq \frac{1}{N}\sum_{i=1}^{N} g(U_i)$$

Ejercicio 3. Calcule mediante un método de Monte Carlo las siguientes integrales:

$$\textbf{i)} \int_{\pi}^{2\pi} \left(\frac{\sin(x)}{x}\right) dx \qquad \textbf{ii)} \int_0^{\infty} \frac{3}{3+x^4}$$

a) Obtenga mediante simulación en computadora el valor de la integral deteniendo la simulación cuando el semi-ancho del intervalo de confianza del 95 % sea justo inferior a 0,001.

c) Indique cuál es el número de simulaciones N_s necesarias en la simulación realizada para lograr la condición pedida y complete con los valores obtenidos la siguiente tabla (usando **4** decimales):

Nº de sim.	$\bar{I}$	S	IC(95 %)
1 000			
5 000			
7 000			
$N_s =$			

a) Tenemos que $1-\alpha = 0.95$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

entonces $P(Z > z_{\alpha/2}) = 0.025$

$$1 - \phi(z_{\alpha/2}) = 0.025$$

$$1 - 0.025 = \phi(z_{\alpha/2})$$

$$\phi^{-1}(0.975) = z_{\alpha/2}$$

$$z_{\alpha/2} = 1.96$$

Entonces en ambos casos el intervalo de confianza sera

$$\left(\bar{X}(n) - 1.96 \frac{S_n}{\sqrt{n}}, \ \bar{X}(n) + 1.96 \frac{S_n}{\sqrt{n}}\right)$$

i)

N^o de sim.	$\bar{I}$	S	IC(95%)
1000	-0.4306	0.2071	(-0.4434, -0.4177)
5000	-0.4293	0.2128	(-0.4352, -0.4234)
7000	-0.4322	0.2095	(-0.4371, -0.4273)
$N_s =$ 170806	-0.4339	0.2109	(-0.4349, -0.4329)

ii)

N^o de sim.	$\bar{I}$	S	IC(95%)
1000	1.3959	1.0082	(1.3334, 1.4583)
5000	1.4835	0.9919	(1.4559, 1.5109)
7000	1.4768	0.9696	(1.4541, 1.4995)
$N_s =$ 3656639	1.4626	0.9756	(1.4616, 1.4636)

Ejercicio 4. Para $U_1, U_2, \ldots$ variables aleatorias uniformemente distribuidas en el intervalo $(0,1)$, se define:

$$N = \text{Mínimo}\left\{n : \sum_{i=1}^{n} U_i > 1\right\}$$

Esto es, N es igual a la cantidad de números aleatorios que deben sumarse para exceder a 1.

a) Observe que $E[N] = e$ por lo cual puede aproximar e con la media muestral $\bar{N}$.

b) Derive una expresión de la varianza del estimador $\bar{N}$ y aproxímela con 1000 simulaciones. Dar su estimador de máxima verosimilitud.

c) Dé el valor obtenido de la varianza muestral de $\bar{N}$ correspondiente a 1000 ejecuciones de la simulación y dar una estimación de e mediante un intervalo de confianza de 95% con longitud a lo sumo 0.025.

a) En el práctico 6 ej.3 demostramos que $E(N)=N$

Como propone el método de monte carlo, con n suficientemente grande si generamos n variables aleatorias uniformes $N_1, N_2, \dots, N_n \sim \mathcal{U}(0,1)$ y $n_1, \dots, n_n$ son observaciones de ellas, es decir $n_1 = N_1(\omega), n_2 = N_2(\omega), \dots, n_n = N_n(\omega)$ entonces:

$$e = E(N) = \frac{1}{n}\sum_{i=1}^{n} n_i = \frac{1}{n}\sum_{i=1}^{n} N_i(\omega) = \bar{N}_n(\omega)$$

Entonces e se puede aproximar por la media muestral $\bar{N}_n$

b) Por lo que vimos en el práctico 3: $P(N=n) = \frac{n-1}{n!}$

$$Var(\bar{N}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} N_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} N_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(N_i)$$

↳ son indep.

y como los N_i tienen igual distribución:

$$Var(\bar{N}) = \frac{1 \cdot \cancel{n} \cdot Var(N_1)}{n^{\cancel{2}}} = \frac{Var(N_1)}{n}$$

$$Var(N_1) = Var(N) = E(N^2) - E(N)^2 = E(N^2) - e^2$$

$$E(N^2) = \sum_{n=1}^{\infty} n^2 \cdot P(N=n) = \sum_{n=1}^{\infty} n^2 \cdot \frac{(n-1)}{n!} = \sum_{n=2}^{\infty} n \cdot \frac{\cancel{(n-1)}}{\cancel{(n-1)}!} = \sum_{n=2}^{\infty} \frac{n}{(n-2)!}$$

$$= \sum_{n=0}^{\infty} \frac{n+2}{n!} = \sum_{n=0}^{\infty} \frac{n}{n!} + 2\sum_{n=0}^{\infty} \frac{1}{n!} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} + 2e$$

↳ $k = n-1$ $\quad n = k+1$ $\quad \begin{matrix} n=1 \\ k=0 \end{matrix}$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} + 2e = e + 2e = 3e$$

$$Var(N_1) = Var(N) = E(N^2) - e^2 = 3e - e^2$$

$$Var(\bar{N}) = \frac{3e - e^2}{n}$$

Este valor es una constante, ya que n lo conocemos, lo definimos para saber cuantos V.A generar, entonces:

- Si $Var(g(x)) = c$ con c constante, entonces
- $\widehat{Var(g(x))} = \hat{c} = c$. Es decir, el EMV de una constante c, es la variable aleatoria que vale constantemente c.

Luego $\widehat{Var(\bar{N})} = \widehat{\frac{3e-e^2}{n}} = \frac{3e-e^2}{n}$

Y el EMV de $Var(\bar{N})$ es la constante $\frac{3e-e^2}{n}$

c) Primero debemos simular 1000 veces $\bar{N}$ y calcular la varianza muestral:

$$S^2 = \frac{1}{1000} \sum_{i=1}^{1000} (N_i - \bar{N}_{1000})^2$$

O $$S^2 = \left(\frac{\sum_{i=1}^{1000} N_i^2}{1000} \right) - (\bar{N}_{1000})^2$$

Luego el intervalo de confianza del 95% con 1000 simulaciones de e es:

$$\left[\bar{N}_{1000} - 1.96 \sqrt{\frac{\hat{\sigma}^2}{1000}} \; ; \; \bar{N}_{1000} + 1.96 \sqrt{\frac{\hat{\sigma}^2}{1000}} \right]$$

Como queremos que la longitud del intervalo sea a lo sumo 0.025 entonces $d = 0.025/2 \cdot 1.96$

Y debemos encontrar la cantidad de simulaciones necesarias

Ejercicio 5. Considere una sucesión de números aleatorios $\{U_i\}_i$ y sea M el primer n tal que la variable U_n es menor que su variable predecesora. Es decir,

$$M = n \quad \text{tal que} \quad U_1 \leq U_2 \leq \cdots \leq U_{n-1} \text{ y } U_n < U_{n-1}$$

a) Justifique que $P(M > n) = 1/n!$, $n \geq 0$.

b) Utilice la identidad

$$E[M] = \sum_{n=0}^{\infty} P(M > n)$$

para mostrar que $E[M] = e$.

c) Utilice el resultado del item anterior para dar un estimador de $E[M]$, calcule el valor de su varianza muestral. Mediante una simulación estime el valor de e deteniéndose cuando la varianza muestral sea menor que 0,01.

d) Dé una estimación de e mediante un intervalo de ancho menor que 0,1 y con una confianza del 95 %

a) $P(M>0) = 1$ ya que tiene que haber al menos 2 variables, luego $P(M>0) = \frac{1}{0!} = 1$

Lo mismo sucede para $P(M>1) = \frac{1}{1!} = 1$.

$P(M>2) = 1 - P(M \leq 2) = 1 - P(M=2) = 1 - P(U_1 > U_2)$

$P(U_2 - U_1 \leq 0) = \int_{-\infty}^{\infty}\int_{-\infty}^{U_1} f_{U_1,U_2}(u_1,u_2)\, du_2\, du_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{U_1} 1 \cdot \mathbb{I}_{(0,1)}(u_1) \cdot \mathbb{I}_{(0,1)}(u_2)\, du_1 du_2$

$= \int_0^1\int_0^{U_1} 1 \cdot du_2\, du_1 = \int_0^1 u_1\, du_1 = \left[\frac{u_1^2}{2}\right]_0^1 = \frac{1}{2}$

$P(M>2) = 1 - \frac{1}{2} = \frac{1}{2} = \frac{1}{2!}$

H.I. Entonces para el caso base queda demostrado, ahora supongamos que para $K \in \mathbb{N}$ con $K > 2$ es cierto que $P(M>K) = \frac{1}{K!}$

Ahora probemos que es cierto para $K+1$

$P(M > K+1) = 1 - P(M \leq K+1)$

$$= 1 - P(M \leq K) - P(M = K+1)$$
$$= 1 - (1 - P(M > K)) - P(M = K+1)$$
$$= P(M > K) - P(M = K+1)$$

$M = K+1$ sucede sólo cuando los primeros K números están ordenados de forma ascendente y $U_{K+1} < U_K$

Primero para calcular la probabilidad de que los primeros K números esten ordenados de forma ascendente, pensemos que tenemos K posiciones.
La probabilidad de que U_1 este en la posición 1 es $1/K$
La probabilidad de que U_2 este en la posición 2 y U_1 en la posición 1 es $1/K-1$
$\vdots$
La probabilidad de que U_K este en la posición K y cada U_i $1 \leq i < K$ este en la posición i es 1
Luego la probabilidad de que los primeros K numeros esten ordenado de forma ascendente es:

$$\frac{1}{K} \cdot \frac{1}{K-1} \cdot \frac{1}{K-2} \cdot \ldots \cdot 1 = \frac{1}{K!}$$

Ahora debemos ver la probabilidad de que $U_K > U_{K+1}$. Si tenemos $K+1$ posiciones y sabemos que U_K esta en la posición $K+1$ entonces U_K puede estar en cualquiera de las K posiciones restantes, es decir $P(U_K > U_{K+1}) = \frac{K}{K+1}$.

Entonces la probabilidad de que tengamos $K+1$ número y que los primeros K números esten ordenados de forma ascendente, con el número $K+1$ en cualquier posición menos en la $K+1$ es:

$$\frac{1}{K!} \cdot \frac{K}{K+1} = \frac{K}{(K+1)!}$$

Luego $P(M > k+1) = P(M>k) - P(M=k+1)$

$$= P(M>k) - \frac{k}{(k+1)!}$$

$$\overset{\{H.I\}}{=} \frac{1}{k!} - \frac{k}{(k+1)!}$$

$$= \frac{1}{k!} - \frac{k}{(k+1)k!}$$

$$= \frac{\cancel{k}+1-\cancel{k}}{k!\,(k+1)}$$

$$= \frac{1}{(k+1)!}$$

Por ende queda demostrado que $P(M>n) = \frac{1}{n!}$ $n \geq 0$

b) $E[M] = \sum_{n=0}^{\infty} P(M>n) = \sum_{n=0}^{\infty} \frac{1}{n!} = e$

c) Para estimar $E[M]=e$ podemos generar n muestras de la variable aleatoria M y calcular el el promedio muestral de dicha muestra, es decir

$$\widehat{E[M]} = \frac{1}{n} \cdot \sum_{i=1}^{n} M_i$$

Y esto funciona ya que $E[\widehat{E[M]}] = E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} M_i\right]$

misma dist. M_i independientes

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} M_i\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} E[M_i] = \frac{1}{\cancel{n}} \cdot \cancel{n} \cdot E[M] = e$$

Ahora veamos la varianza

$$Var(\widehat{E[M]}) = Var\left(\frac{1}{n}\cdot\sum_{i=1}^{n} M_i\right) = \frac{1}{n^2}\cdot Var\left(\sum_{i=1}^{n} M_i\right) \overset{\text{misma dist.}}{=} \frac{1}{n^2}\cdot\sum_{i=1}^{n} Var(M_i)$$

var. indep.

$$= \frac{1}{n^2}\cdot n Var(M) = \frac{Var(M)}{n}$$

$$Var(M) = E[M^2] - E[M]^2$$
$$= E[M^2] - e^2$$

$$E[M^2] = \sum_{n=0}^{\infty} n^2 \cdot P(M=n) = \sum_{n=0}^{\infty} n^2 \cdot \frac{1}{n!} = \sum_{n=1}^{\infty} \frac{n}{(n-1)!}$$

$k = n-1 \qquad n=1$

$n = k+1 \qquad k=0$

$$= \sum_{k=0}^{\infty} \frac{k+1}{k!} = \sum_{k=0}^{\infty} \frac{k}{k!} + \sum_{k=0}^{\infty} \frac{1}{k!} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} + e$$

$$= 2e$$

$$Var(M) = 2e - e^2$$

$$Var(\widehat{E[M]}) = \frac{2e-e^2}{n}$$

d) El IC del 95% para e es

$$\left[\overline{X}(M) - 1.96\sqrt{\frac{\sigma^2}{n}} \; ; \; \overline{X}(M) + 1.96\sqrt{\frac{\sigma^2}{n}}\right]$$

Además queremos que el ancho del intervalo sea menor a 0.1

En particular como ya sabemos que $\sigma^2 = 2e - e^2$ entonces

$$IC = \left[\bar{X}(M) - 1{,}96 \cdot \sqrt{\frac{2e-e^2}{n}} \; ; \; \bar{X}(M) + 1{,}96 \sqrt{\frac{2e-e^2}{n}} \right]$$

Ejercicio 6. Estime π sorteando puntos uniformemente distribuidos en el cuadrado cuyos vértices son: $(1,1), (-1,1), (-1,-1), (1,-1)$, y contabilizando la fracción que cae dentro del círculo inscrito de radio 1.

a) Utilice un algoritmo para estimar la proporción de puntos que caen dentro del círculo y deténgase cuando la desviación estandar muestral del estimador sea menor que 0,01.

b) Obtenga un intervalo de ancho menor que 0,1, el cual contenga a π con el 95 % de confianza. ¿Cuántas ejecuciones son necesarias?

a) Tenemos que generar 2 variables aleatorias $X \sim \mathcal{U}(-1,1)$ $Y \sim \mathcal{U}(-1,1)$ luego tenemos que ver si $\underbrace{\sqrt{x^2+y^2}}_{\text{distancia al origen}} <= 1$ ya que el radio $= 1$.

Con esto tenemos una proporción de los puntos que caen dentro del círculo.

Si llamamos X a la variable aleatoria Bernoulli que vale 1 si el punto generado cae dentro del círculo y 0 si no, podemos estimar la proporción usando la media muestral de X, es decir $\hat{p}(n) = \bar{X}(n)$ con n el tamaño de la muestra.

Luego como X es una Bernoulli:

$$\sigma^2 = p \cdot (1-p)$$
$$\hat{\sigma}^2 = \hat{p}(n) \cdot (1-\hat{p}(n))$$
$$\hat{\sigma}^2 = \bar{X}(n) \cdot (1-\bar{X}(n))$$

Entonces la varianza muestral es $Var(\bar{X}(n)) = \dfrac{\bar{X}(n) \cdot (1-\bar{X}(n))}{n}$

b) Para estimar el valor de π usamos la proporción estimada en el ejercicio a. La proporción, que es igual a la cantidad de puntos que caen dentro del círculo, nos ayuda a estimar el área del círculo, ya que dividimos la cantidad de puntos que caen dentro del área del círculo por la cantidad de puntos generados (es decir dentro del área del cuadrado). Ahora tenemos la fración del área del cuadrado que esta ocupada por el círculo.

Recordemos que el círculo tiene radio 1, entonces el área del círculo es $\pi \cdot r^2 = \pi \cdot 1^2 = \pi$. El cuadrado tiene lados de longitud 2 $(-1,1)$, entonces el área del cuadrado es $2^2 = 4$.

Entonces la fración del área del cuadrado que está ocupada por el círculo es $\pi/4$, entonces para estimar el valor de π debemos calcular $4 \times \frac{(\text{cantidad de puntos en el círculo})}{(\text{cantidad de puntos totales})}$