ESTADÍSTICA (Q)

SIMULACIÓN - TEOREMA CENTRAL DEL L?MITE - CLASE 10

Antes de empezar, los invitamos a visitar el sitio Point of Significance, una publicación de Nature dedicada a la divulgaci?n de la estad?stica dentro de las ciencias naturales. En particular, los invitamos a que miren el trabajo Importance of being uncertain, considerando que vamos a querer replicar parte de los resultados presentados en la Figura 3.

A lo largo de esta Gu?a estudiaremos emp?ricamente la distribución del promedio de variables aleatorias independientes e idénticamente distribuidas. A través de los histogramas correspondientes, analizaremos el comportamiento de la distribuci?n del promedio a medida que aumentamos n, la cantidad de variables a promediar.

Para ello generaremos un conjunto de n datos con una distribución dada y luego calcularemos su promedio. Replicaremos ésto mil veces, es decir, generaremos $Nrep = 1000$ realizaciones de la variable aleatoria $\overline{X}_n$, para diferentes valores de n. Observemos que, en principio, desconocemos la distribución de $\overline{X}_n$. Utilizando las $Nrep = 1000$ realizaciones del promedio realizaremos un histograma de los promedios generados para obtener una aproximación de la densidad o la función de probabilidad de $\overline{X}_n$.

1 Teorema Central del L?mite: El Teorema

Antes de empezar con nuestras simulaciones, recordemos el Teorema Central del L?mite. Sea $(W_i)_{i\geq 1}$ una sucesión de v.a.i.i.d. con $\mathbb{E}(W_i) = \mu$ y $\mathbb{V}(X_i) = \sigma^2$.

Una de las maneras de ver el TCL en t?rminos del promedio es la siguiente:

$$\overline{W}_n \overset{a}{\approx} \mathcal{N}(\mu, \sqrt{\sigma^2/n}) \ . \tag{1}$$

Si estandarizamos al promedio, obtenemos esta otra posible presentaci?n:

$$\frac{\overline{W}_n - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\approx} \mathcal{N}(0,1) \ . \tag{2}$$

2 Teorema Central del L?mite: Simulaciones

2.1 Distribuci?n Bernoulli

Sean $X_i \sim \mathcal{B}e(p)$ siendo $p = 0,2$. Utilizaremos X para denotar de manera gen?rica una variable con esta misma distribuci?n.

1. ¿Cu?nto vale $\mathbb{E}(X)$? ¿Cu?nto vale $\mathbb{V}(X)$?

2. Guardar en el vector ber_N_infty= 10000 datos correspondientes a 10.000 realizaciones $\mathcal{B}e(p)$ con $p = 0,2$ y calcular la frecuencia relativa de cada posible valor. Explorar el comando `table()`.

   La distribuci?n emp?rica del promedio

3. Guardar en el vector `promedios_bernoullies`, $Nrep = 1000$ promedios utilizando $n = 5$ datos con distribuci?n $\mathcal{B}e(p)$ siendo $p = 0,2$. Recordar el comando `rbinom()`, teniendo precauci?n con los par?metros que utiliza esta funci?n (ver `help(rbinom)`).

4. Realizar un histograma con los promedios guardados en `promedios_bernoullies`.

5. Repetir los ?tems 2 y 3 con $n = 30$ y con $n = 100$. Comparar los tres histogramas obtenidos, notar la diferencia en las escalas.

   La distribuci?n emp?rica del Promedio *estandarizado*

6. Realizar ahora histogramas de `promedios_bernoullies_est` donde a los valores guardados en `promedios_bernoullies` se les resta μ y se los divide por $\sqrt{\sigma^2/n}$. Utilizar $n \in \{5, 30, 100\}$, ¿Qu? se observa?

7. Repetir todos los ?tems anteriores con $p = 0.5$.

2.2 Distribuci?n Uniforme

Sean $X_i \sim \mathcal{U}(a, b)$ siendo $a = 67$, $b = 73$. Utilizaremos X para denotar de manera gen?rica una variable con esta misma distribuci?n.

1. ¿Cu?nto vale $\mathbb{E}(X)$? ¿Cu?nto vale $\mathbb{V}(X)$?

2. Guardar en el vector unif_N_infty= 10000 datos correspondientes a 10.000 realizaciones $\mathcal{U}(a, b)$ siendo $a = 67$, $b = 73$ y realizar un histograma.

   La distribuci?n emp?rica del promedio

3. Guardar en el vector `promedios_uniformes`, $Nrep = 1000$ promedios utilizando $n = 5$ datos con distribuci?n $\mathcal{U}(a, b)$ siendo $a = 67$, $b = 73$. Recordar el comando `runif()`.

4. Realizar un histograma con los promedios guardados en `promedios_uniformes`.

5. Repetir los ?tems 2 y 3 con $n = 30$ y con $n = 100$. Comparar los tres histogramas obtenidos, notar la diferencia en las escalas.

La distribuci?n emp?rica del Promedio *estandarizado*

6. Realizar ahora histogramas de `promedios_uniformes_est` donde a los valores guardados en `promedios_uniformes` se les resta μ y se los divide por $\sqrt{\sigma^2/n}$. Utilizar $n \in \{5, 30, 100\}$, ¿Qu? se observa?

2.3 Distribuci?n Exponencial

Sean $X_i \sim \mathcal{E}(\lambda)$ siendo $\lambda = 1/10$. Utilizaremos X para denotar de manera gen?rica una variable con esta misma distribuci?n.

1. ¿Cu?nto vale $\mathbb{E}(X)$? ¿Cu?nto vale $\mathbb{V}(X)$?

2. Guardar en el vector exp_N_infty= 10000 datos con distribuci?n $\mathcal{E}(\lambda)$ siendo $\lambda = 0,1$ y realizar un histograma de ellos.

La distribuci?n emp?rica del promedio

3. Guardar en el vector `promedios_exponenciales`, $Nrep = 1000$ promedios utilizando $n = 5$ datos con distribuci?n $\mathcal{E}(\lambda)$ siendo $\lambda = 0,1$. Recordar el comando `rexp()`.

4. Realizar un histograma con los promedios guardados en `promedios_exponenciales`.

5. Repetir los ?tems 2 y 3 con $n = 30$ y con $n = 100$. Comparar los tres histogramas obtenidos, notar la diferencia en las escalas.

La distribuci?n emp?rica del Promedio *estandarizado*

6. Realizar ahora histogramas de `promedios_exponenciales_est` donde a los valores guardados en `promedios_exponenciales` se les resta μ y se los divide por $\sqrt{\sigma^2/n}$. Utilizar $n \in \{5, 30, 100\}$, ¿Qu? se observa?

7. Realizar ahora histogramas de `promedios_exponenciales` con $n \in \{5, 30, 100\}$ donde a los valores guardados en `promedios_exponenciales` se les resta μ y se los divide por $\sqrt{\sigma^2/n}$. ¿Qu? se observa?

3 Simulaciones de promedios de variables aleatorias Normales

Sean $X_i \sim \mathcal{N}(\mu, \sigma^2)$ siendo $\mu = 70$ y $\sigma^2 = 1.2$. Utilizaremos X para denotar de manera gen?rica una variable con esta misma distribuci?n.

1. ¿Cu?nto vale $\mathbb{E}(X)$? ¿Cu?nto vale $\mathbb{V}(X)$?

2. Guardar en el vector norm_N_infty= 10000 datos con distribuci?n $\mathcal{N}(\mu, \sigma^2)$ siendo $\mu = 70$ y $\sigma^2 = 1.2$ y realizar un histograma de ellos.

   La distribuci?n emp?rica del promedio

3. Guardar en el vector `promedios_normales`, $Nrep = 1000$ promedios utilizando $n = 5$ datos con distribuci?n $\mathcal{N}(\mu, \sigma^2)$ siendo $\mu = 70$ y $\sigma^2 = 1.2$. Recordar el comando `rnorm()`.

4. Realizar un histograma con los promedios guardados en `promedios_normales`.

5. Repetir los ?tems 2 y 3 con $n = 30$ y con $n = 100$. Comparar los tres histogramas obtenidos, notar la diferencia en las escalas.

   La distribuci?n emp?rica del Promedio *estandarizado*

6. Realizar ahora histogramas de `promedios_normales_est` donde a los valores guardados en `promedios_normales` se les resta μ y se los divide por $\sqrt{\sigma^2/n}$. Utilizar $n \in \{5, 30, 100\}$, ¿Qu? se observa?