Trabajo Práctico 2 en R
(¡ahora en formato presentación!)

Ley de los Grandes Números y Teorema Central del Límite

Cosas a tener en cuenta para la correcta entrega del TP

- Deben subir las diapositivas (formato `.pdf`) y el código (formato `.R`, `.Rmd` o `.qmd`) **24h antes de la fecha de exposición**.
- Para generar los gráficos elijan como semilla su número de grupo. Por ejemplo, si son el grupo 4, al comienzo del archivo que usaron para generar los gráficos deben poner `set.seed(4)`.
- La exposición debe durar <u>máximo</u> 10 minutos. Sean concisos y claros.
- Siéntanse libres de responder las preguntas como les parezca más adecuado. Hagan tablas, gráficos o lo que necesiten para complementar las discusiones.

1 – Variables aleatorias con distribución uniforme

Sea $X \sim U(0, 18)$ una variable distribuida uniformemente entre 0 y 18

(1.a) Calculen $\mathbb{E}(X)$ y $\mathbb{V}(X)$. Pueden utilizar las fórmulas conocidas para dicha distribución.

Muestren la cuenta en la presentación.

(1.b) Generen una función `X_dist(R)` que devuelva un vector con R realizaciones de X.

No es necesario que muestren el código en la presentación pero sí que esté en el archivo de códigos.

(1.c) Hagan dos histogramas de X, tomando $R \in \{100, 10^4\}$ realizaciones y 30 bines.
¿Qué distribución ven? Discutan qué efecto tiene variar R en este contexto.

(1.d) Utilizando la función `X_dist(R)`, generen un vector `muchas_X` que contenga $R = 10^4$ realizaciones de X. Luego, para cada n entre 1 y 10^4 calculen el promedio de los n primeros valores de `muchas_X`. Grafiquen esos promedios en función de n e incluyan en el gráfico una recta horizontal a la altura de $\mathbb{E}(X)$.

Discutan el resultado en el contexto de la ley de los grandes números.

2 – Variables aleatorias con distribución exponencial
Sea $Y \sim \mathcal{E}(1/9)$ una variable con distribución exponencial de parámetro $\lambda = 1/9$.

(2.*) Repitan lo realizado en el punto anterior (salvo el ítem **(1.d)**) pero para Y en vez de X. Si así lo desean, y les parece más ordenado, pueden exponer este punto y el anterior de forma intercalada en la presentación.

3 – Promedios de variables aleatorias
Es posible construir variables aleatorias como combinaciones de otras variables aleatorias, como la variable aleatoria "suma del resultado de dos dados" o "división entre la cantidad de gente que compra algo en una web y la cantidad de gente que la visita por día". En esta sección vamos a trabajar con la variable aleatoria "hacer quince repeticiones de X, sumar los resultados y dividir todo por quince"; también conocido como el promedio de quince valores independientes de X:

$$\overline{X}_{15} = \frac{1}{15}\big(X_1 + X_2 + \cdots + X_{15}\big) = \frac{1}{15}\sum_{i=1}^{15} X_i;$$

Comentario: También podrían hacer esto para quince valores independientes de Y:

$$\overline{Y}_{15} = \frac{1}{15}\big(Y_1 + Y_2 + \cdots + Y_{15}\big) = \frac{1}{15}\sum_{i=1}^{15} Y_i.$$

(3.a) Utilizando los resultados obtenidos en **(1.a)** y las propiedades de la esperanza y la varianza, calculen $\mathbb{E}(\overline{X}_{15})$, $\mathbb{V}(\overline{X}_{15})$.
Muestren la cuenta en la presentación.
¿Qué efecto tiene sobre la esperanza el tomar un promedio?
¿Qué efecto tiene sobre la varianza el tomar un promedio?

(3.b) Generen una función `Xn_dist(n, R)` que devuelva un vector con R realizaciones del promedio de n valores de X.

No es necesario que muestren el código en la presentación pero sí que esté en el archivo de códigos.

(3.c) Hagan un histograma de $\overline{X}_{15}$, tomando $R = 10^4$ ¿Qué distribución ven?

4 – TCL: Teorema Central del Límite

Lo observado en **(3.c)** es consecuencia del Teorema Central del Límite. Este teorema establece que, en condiciones bastante generales, al tomar una muestra aleatoria de una población con cualquier distribución, la media de esta muestra tendrá distribución aproximadamente normal, siempre que el tamaño de la muestra sea lo suficientemente grande. Es decir, que

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{para } n \text{ grande.}$$

En esta sección vamos a trabajar con las mismas variables aleatorias que introdujimos en las secciones anteriores:

$$X \sim U(0,18), \qquad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \approx \mathcal{N}\left(\mu_X, \frac{\sigma_X^2}{n}\right) \quad \text{con} \quad \mu_X = \mathbb{E}(X) \ \text{ y } \ \sigma_X^2 = \mathbb{V}(X),$$

$$Y \sim \mathcal{E}(1/9), \qquad \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \approx \mathcal{N}\left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \quad \text{con} \quad \mu_Y = \mathbb{E}(Y) \ \text{ y } \ \sigma_Y^2 = \mathbb{V}(Y),$$

donde ahora para el promedio tomamos n como una variable libre que vamos a ir cambiando, en vez de fijar $n = 15$ como hicimos en la sección anterior.

(4.a) Muestren que se cumple el TCL para un caso particular. Para esto realicen un histograma para X y otro para $\overline{X}_{40}$, con $R = 10^6$ realizaciones en ambos casos. Superpongan la curva correspondiente a la distribución teórica (exacta o aproximada) de cada variable aleatoria. Hagan lo mismo para Y y $\overline{Y}_{40}$.

(4.b) Grafiquen diez histogramas como un arreglo de 5×2 paneles (filas$\times$columnas). En la primera fila, grafiquen los histogramas de $\overline{X}_n$ para $n = 1$ y $R \in \{10^2, 10^6\}$ (un panel por cada R). Para las siguientes filas repitan este procedimiento con $n \in \{2, 5, 15, 40\}$. ¿Qué cambia y qué no al variar n? ¿Qué cambia y que no al variar R?
Discutan el resultado detalladamente.

(4.c) Repitan lo realizado en el punto anterior pero para $\overline{Y}_n$. Discutan la diferencia entre la convergencia de $\overline{X}_n$ y $\overline{Y}_n$ a medida que aumenta n.