1. El objetivo de este ejercicio es utilizar la Ley de los Grandes Números para aproximar integrales de funciones continuas en intervalos finitos. Procuraremos aproximar la probabilidad de que una variable $Z \sim \mathcal{N}(0,1)$ tome valores en un intervalo $[a,b]$. Es decir, queremos aproximar numéricamente la siguiente integral

$$\int_a^b \frac{1}{\sqrt{2\pi}}\, exp^{-\frac{x^2}{2}}\, dx\ . \tag{1}$$

Para ello, consideremos la siguiente propuesta: tomamos $U_1, U_2, \ldots U_n$ variables i.i.d. $U_i \sim \mathcal{U}[a,b]$ independientes y calculamos

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\, exp^{-\frac{U_i^2}{2}}}{n}$$

(a) Aproximamos la integral (1) mediante

$$(b-a)\,\frac{\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\, exp^{-\frac{U_i^2}{2}}}{n}\ ,$$

para los siguientes valores de a y b, y para $n = 100$, 1000 y 50000.

i. $a = -1.96$ y $b = 1.96$.
ii. $a = -2$ y $b = 1$.
iii. $a = 0$ y $b = 2.34$.

(b) Otro modo de aproximar la probabilidad de que una variable $Z \sim \mathcal{N}(0,1)$ tome valores en un intervalo $[a,b]$. Sería generar variables $Z_1, Z_2, \ldots Z_n$ i.i.d. $Z_i \sim \mathcal{N}[0,1]$ y calcular

$$\frac{1}{n}\sum_{i=1}^{n} I_{(a,b)}(Z_i).$$

Explique por que es razonable este métdodo. Y aproxime la probabilidad para los siguientes valores de a y b, y para $n = 100$, 1000 y 50000.

i. $a = -1.96$ y $b = 1.96$.
ii. $a = -2$ y $b = 1$.
iii. $a = 0$ y $b = 2.34$.

(c) Comparar los resultados obtenidos en a) y b) con aquellos que calcularía utilizando con R la función Φ.

2. Cuando se realiza una encuesta con preguntas delicadas, donde la gente tiene cierta resistencia a responder, se suele usar una método indirecto para hacer la pregunta. Supongamos que estamos interesados en conocer (o estimar) la proporción de mujeres que se realizaron un aborto. Un modo de realizar la encuesta sería el siguiente: a cada mujer le damos un dado y una moneda y le decimos que tire el dado. Si el resultado del dado es 3, 4, 5, o 6 la mujer responde la verdad y si sale 1 o 2 tira la moneda; si sale cara responde SI y si sale ceca responde NO.

   Obviamente, el encuestador no ve el resultado del dado ni el de la moneda y por lo tanto no sabría si la respuesta del encuestado es la verdad o es producto del azar.

   A partir de este mecanismo quisieramos poder estimar la verdadera proporción de mujeres que se hicieron un aborto que llamaremos p, es decir p =Probabilidad de que una mujer elegida al azar se haya realizado un aborto.

   (a) Calcular (en función de p) la probabilidad de que una persona elegida al azar responda afirmativamente a la pregunta ($P(SI)$). Ayuda: condicione al resultado del dado.

   (b) A partir de encuestar a n personas con este mecanismo, como estimaría la probabilidad de que alguien responda que SI, es decir $P(SI)$? Mas precisamente, si pensamos en n variables i.i.d $W_1, \ldots, W_n$ donde $W_i = 1$ si la mujer responde SI y 0 en caso contario, que cuenta deberá hacer para estimar la proporción de gente que responde SI, P(SI).

   (c) Usando a) y b), proponga una cuenta para estimar p, la probabilidad de que una mujer elegida al azar se haya realizado un aborto. Llamemos $\hat{p}$ a esta cuenta.

   (d) Para convencernos que este mecanismo resulta efectivo, realice la siguiente simulación. Supongamos que el verdadero valor de $p = 0.1$ es decir la verdadera proporción de mujeres que se realizó un aborto. Simule $n = 100$ respuestas, $W_1, \ldots, W_n$ según el mecánismo de respuesta propuesto y compare con la verdadera proporción p. Repita esto $Nrep = 1000$ veces.

   (e) Si $p = 0.1$, ¿ qué distribución tiene W_i? Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \hat{p} < 0.2)$.

   (f) Supongamos como en el ejercicio anterior que $p = 0.1$ pero hacemos la encuenta directa y nadie miente. En ese caso, estimaríamos a p simplemente contando el porcentaje de gente que contesta si. Es decir, tendríamos $X_1, \ldots, X_n$ variables aleatorias i.i.d $Bi(1, p)$ y estimaríamos p con $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \bar{X}_n < 0.2)$.

   (g) Explique la relación o las diferencias entre los dos item anteriores.

3. En este ejercicio estudiaremos la distribución del promedio $\bar{X}_n$ de variables $X_1, \ldots, X_n$ (i.i.d.) pero con distribución asimétrica. Consideremos X con distribución LogNormal(μ, σ^2), es decir

su densidad es

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log x - \mu)^2\right), \qquad x > 0.$$

En tal caso, la esperanza y varianza de X están dadas por

$$\mathrm{E}(X) = e^{\mu+\sigma^2/2} \quad \text{y} \quad \mathrm{var}(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2},$$

repectivamente. El nombre de la distributicón proviene del siguiente hecho:

$$X \sim \text{LogNormal}(\mu, \sigma^2) \quad \text{si y solo si} \quad \log(X) \sim \mathcal{N}(\mu, \sigma^2).$$

(a) Consideremos una variable X_1 con distribución LogNormal(0,2). Grafiquemos su densidad. Indique el valor de $\mathrm{E}(X_1)$ y el de $\mathrm{var}(X_1)$.

(b) Generaremos datos correspondientes a una muestra $X_1, \ldots, X_n$ con distribución LogNormal(0,2) y computamos

$$T_n = \frac{\bar{X}_n - E(X_1)}{\sqrt{Var(\bar{X}_n)}} = \frac{\bar{X}_n - E(X_1)}{\sqrt{Var(X_1)/n}}\ .$$

Replicamos $Nrep = 1000$ veces, para diferentes valores de n: $n = 30, 100, 500, 1000$, obteniendo 4 conjuntos de datos. Cada uno de ellos contienen los $Nrep = 1000$ valores de T_n obtenidos para $n = 30, 100, 500, 1000$, respectivamente.

   i. Realice un histograma para conjunto de T_n obtenidos. ¿Qué características tienen estos histogramas? ¿Qué se observa? Notemos que para poder comparar los histogramas de los distintos conjuntos de datos será necesario representarlos en la misma escala, tanto para el eje horizontal como para el vertical.

   ii. Realice también los boxplots para comparar los distintos conjuntos de datos en un mismo gráfico. ¿Qué observa?