1. **Comprobación empírica de la LGN**: caclulalemos promedios utilizando cada vez más datos observaremos el comportamiento del promedio a medida que la cantidad de datos aumenta.

   (a) Considere una distribución $\mathcal{U}(0,1)$. Indique cuál es el valor verdadero de la media μ.

   (b) Genere una muestra de tamaño $N = 1000$ y para cada k entre 1 y 1000 calcule el promedio de las primeras k observaciones. Realice un scatterplot de k vs. promedio. Incluya una linea horizontal en el valor de y correspondiente a la verdadera media μ.

   (c) Repetimos b) comenzando con otra semilla y superpongamos el nuevo gráfico utilizando un color diferente. Comparamos los gráficos obtenidos. ¿ Qué puede concluir?

   (d) Genere ahora datos con una distribución de Cauchy. Recuerde que la distribución Cauchy coincide con la distribución t de Student con un grado de libertad, y su dendsidad está dada por

   $$f_X(x) = \frac{1}{\pi(1+x^2)} \,.$$

   Observe que es una densidad simétrica alrededor del cero, con colas que acumulan más probabilidad que la normal estándar, y que no tiene esperanza ni varianza finitas. Repita el item b) generando ahora variables con distribución Cauchy, utilizando en R el comando rt(n, df=1). Repita utilizando diferentes semillas. Comente los gráficos obtenidos.

2. En este ejercicio estudiaremos de manera empírica la distribución del promedio $\bar{X}_n$ de variables $X_1, \ldots, X_n$ independientes e idénticamente distribuidas (i.i.d.). Para ello generaremos datos y realizaremos histogramas de promedios $\bar{X}_n$. Es decir, trataremos de interpretar mediante una simulación resultados de la Ley de los Grandes Números y el Teorema Central del Límite.

   (a) Caso $n = 1$: la variable coincide con el promedio. Genere $Nrep = 1000$ datos correspondientes a $X \sim \mathcal{U}(0; 1)$ y luego haga un histograma con los datos generados.
   ¿A qué densidad se parece el histograma obtenido?

   (b) Caso $n = 2$: Genere $n = 2$ datos (independientes) correspondientes a una variable aleatorias con distribución $\mathcal{U}(0; 1)$ y compute el promedio. Replique $Nrep = 1000$ veces y realice un histograma con los $Nrep = 1000$ promedios obtenidos. ¿Qué características tiene este histograma?

   (c) Repita el item anterior utilizando $n = 5$, $n = 30$, $n = 500$, $n = 1200$ y guarde el valor de los $Nrep = 1000$ promedios obtenidos para cada valor de n.

   (d) Realice un histograma para cada conjunto de promedios. Observe que un boxplot es más cómodo para comparar distintos conjuntos de datos.

(e) Calculamos media y el desvio muestral para cada conjunto de datos. ¿A qué valores teóricos deberían parecerse?

(f) El Teorema Central del Límite establece que la distribución de

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

se aproxima a la distribución normal de la normal estándar, cuando n es suficientemente grande.

Transforme los datos obtenidos en el item c) y estudie la distribución de cada conjunto de datos transformado mediante histogramas y boxplots.

(g) Repitamos el ítem e) generando ahora variables con distribución Cauchy C(0; 1). Recuerde de C(0;1) coincide con Student con 1 grado de libertad (rt(n,df=1) es para generar en R). Comparemos los resultados obtenidos. ¿Se puede realizar el ítem f)? ¿Por qué?

3. Intervalos de Confianza: *Interval that contains an unknown quantity with a given frequency- All of Statistics, Wasserman.*

   (a) Implemente una función **intervalo.mu.asin** que tenga por input el nivel $1-\alpha$ y un conjunto de datos $x_1, \ldots, x_n$, provenientes de una muestra de X, y devuelva el intervalo de confianza de nivel asintótico $1-\alpha$ para $\mu = \mathbb{E}(X)$.

   Nivel de Cubrimiento empírico: El nivel de cubrimiento empírico (de un procedimiento) se define como la proporción de veces que los intervalos (construidos con el procedimiento) utilizando datos simulados contiene a μ (o θ), en cierta cantidad de $Nrep$ replicaciones.

   (b) **Simulación bajo normalidad.** Genere variables con distribución normal de media $\mu = 0$ y $\sigma = 0.1, 1, 10$. Calcule el cubrimiento empírico del intervalo de confianza asintótico de nivel 1-α, definido en **intervalo.mu.asin**, para $\alpha = 0.05$, $n = 5$, $n = 10$, $n = 30$, $n = 50$, $n = 100$, $n = 1000$, utilizando $Nrep = 1000$ replicaciones, y complete la siguiente tabla. En cada caso, calcule el promedio de las longitudes en las $Nrep = 1000$ replicaciones e inlcuya el valor en la tabla (long). Indique a que valor debe aproximarse el nivel de cubrimiento empírico. Comente los resultados obtenidos. En base a los resultados, ¿parece razonable este procedimiento para $n = 5$? ¿Conoce algún procedimiento alternativo?

4. Algo de Bootstrap: Implemente una función **intervalo.mediana.boot** que tenga por input el nivel $1-\alpha$ y un conjunto de datos $x_1, \ldots, x_n$, provenientes de una muestra de X, y devuelva el intervalo de confianza asintótico $1-\alpha$ para la mediana de X.

   Para ello, asuma que la mediana muestral es asintóticamente normal y estime el desvío del estimador mediante remuestreo con $Nboot = 100$.

Modelo	Normales con media $\mu = 0$. Nivelo nominal=0.9					
	n=5	n=10	n=30	n=50	n=100	n=1000
$\sigma = 0.1$						
$\sigma = 1$			cub (long)			
$\sigma = 10$						

5. Simulación bajo Cauchy. Genere variables con distribución Cauchy $\mathcal{C}(0,1)$. Calcule el cubrimiento empírico del intervalo de confianza bootstrap para la mediana, definido en intervalo.mediana.boot, para $\alpha = 0.05$, $n = 30$, $n = 50$, $n = 100$ y $n = 1000$, utilizando $Nrep = 1000$ replicaciones, y complete la siguiente tabla. En cada caso, calcule el promedio de las longitudes en las $Nrep = 1000$ replicaciones y el cubrimiento empírico. Indique a que valor debe aproximarse el nivel de cubrimiento empírico. Comente los resultados obtenidos.

	n=30	n=50	n=100	n=1000
longitud				
cubrimiento				

6. Tarea para el hogar. 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 $Nrep = 10000$ veces, para n: $n = 30, 100, 500, 1000$, obteniendo 4 conjuntos de datos.

(c) Realice un histograma para cada conjunto de datos y comente sus principales características.