- **Fecha máxima de entrega:** Jueves 25 de febrero @ 19:00 h

Determinístico

I) Calcule $n!$; la correspondiente **aproximación de Stirling** $\big(S(n)\big) = n! \approx \sqrt{2\pi n}n^n e^{-n}$; la diferencia entre ellas $\big(D(n) = n! - S(n)\big)$; y la diferencia relativa $\big(DR(n) = D(n)/n!\big)$ para $n = 1, 2, \ldots, 12$.

Tabla 1: Error de aproximación $D(n)$ de la aproximación de *Stirling* $S(n)$ al factorial $n!$.

n	$n!$	$S(n)$	$D(n)$	$D(n)/n!$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				

Simulación

Sea X una variable aleatoria (*v. a.*) definida sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Para $w \in \Omega$, $X(w) = x \in \mathbb{R}$ se dice una **realización** de la variable X. Si uno construye un **histograma** para *muchas* realizaciones x, el diagrama describe la distribución de X aproximando f, la función de densidad de probabilidad (*f. d. p.*) de X. Mientras mayor sea el número de realizaciones de X mejor será la aproximación.

Considere la *v. a.* X con *f. d. p.* f. Simule N realizaciones de la variable y construya el correspondiente histograma. Por ejemplo, para el caso de $X \sim \text{Gamma}(\alpha = 2, \beta = 1)$ con $N = 20,000$ realizaciones, el código usando R sería el siguiente y da lugar a la Figura 1.

```
> set.seed(20240905)
N <- 20000
x <- rgamma(N, shape=2, scale=1)
out <- hist(x, xlab="observaciones", ylab="frecuencia", main="Distribución Gamma")
```

II) Ejecute el código anterior. Note que el objeto `out` es de clase “histogram”. Investigue qué componentes tiene la lista `out`.

[**Nota**: En este inciso no debe reportar nada pero sí realizar lo indicado.]

Con base en la simulación anterior determine $p_N = \mathbb{P}_N\big(3 < X \leq 4\big)$ y compárelo con $p = \mathbb{P}\big(3 < X \leq 4\big)$, la probabilidad teórica calculada con base en la distribución de X,

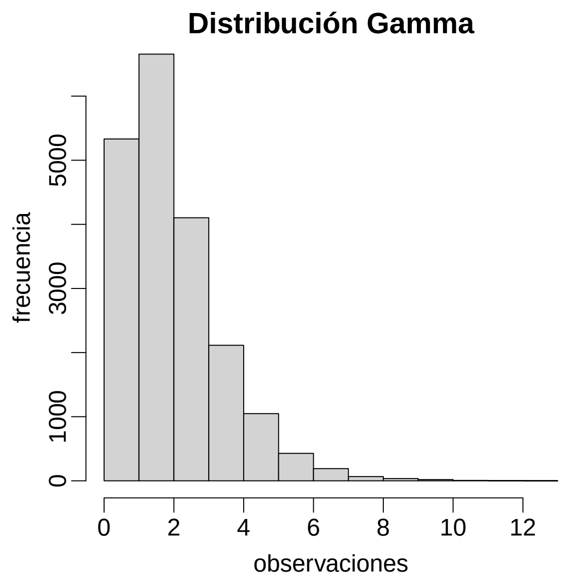

Figura 1: Histograma de 20000 realizaciones de una distribución gamma de parámetros $\alpha = 2$ y $\beta = 1$.

Gamma($\alpha = 2, \beta = 1$). La diferencia entre ellas es el **error de estimación**, $\epsilon = p_N - p = 0.1079 - 0.1076 = -0.0003$.

Teorema Central del Límite

Suponga X *v. a.* con *f. d. p.* f, media $\mu = \mathbb{E}[X]$ y varianza σ^2 ($< \infty$). Sea $\{X_1, \ldots, X_n\}$ una *m. a.* de tamaño n. La **media muestral** o **promedio aritmético** se define como

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

donde el subíndice n hace explícito el tamaño de la muestra. El **Teorema Central del Límite** (*TCL*) afirma que

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim \mathrm{N}(0,1)$$

Considere $\{x_{i1}, \ldots, x_{in}\}$ la i-ésima muestra observada (simulada) de la población X y sea $\bar{x}_i$ la correspondiente media muestral, $i = 1, \ldots, N$. Luego, para N grande, se puede construir el histograma para $\{\bar{x}_i\}$ que aproxime la distribución de $\bar{X}_n$. Por el *TCL* se espera que el histograma se parezca cada vez más a la *campana normal* conforme crece n.

Si como en el ejemplo, $X \sim$ Gamma($\alpha = 2, \beta = 1$) entonces $\mu = \alpha\beta = 2$ y $\sigma^2 = \alpha\beta^2 = 2$, por lo que se sigue del *TCL* que para n *grande*

$$Z_n := \frac{\bar{X}_n - 2}{\sqrt{2/n}} \dot{\sim} \mathrm{N}(0,1), \quad \text{equivalentemente} \quad \bar{X}_n \dot{\sim} \mathrm{N}(2, 2/n)$$

donde $\dot{\sim}$ indica que la distribución es aproximada.

¿Qué tan buena es la aproximación? La figura 2 muestra el histograma de 20000 simulaciones de promedios estandarizados de la distribución Gamma(2,1) de muestras de tamaño 25 sobrepuesto sobre el correspondiente diagrama de barras, que no histograma, de la distribución normal estándar, construido con base en la distribución teórica. La diferencia en las distribuciones es aparente. En este ejemplo, el error de aproximación es de 0.0808, la suma de los errores absolutos de cada uno de los intervalos del histograma. Es muy probable que no se reproduzca este error pues fue calculado con distinto código y sin considerar la misma *semilla aleatoria*.

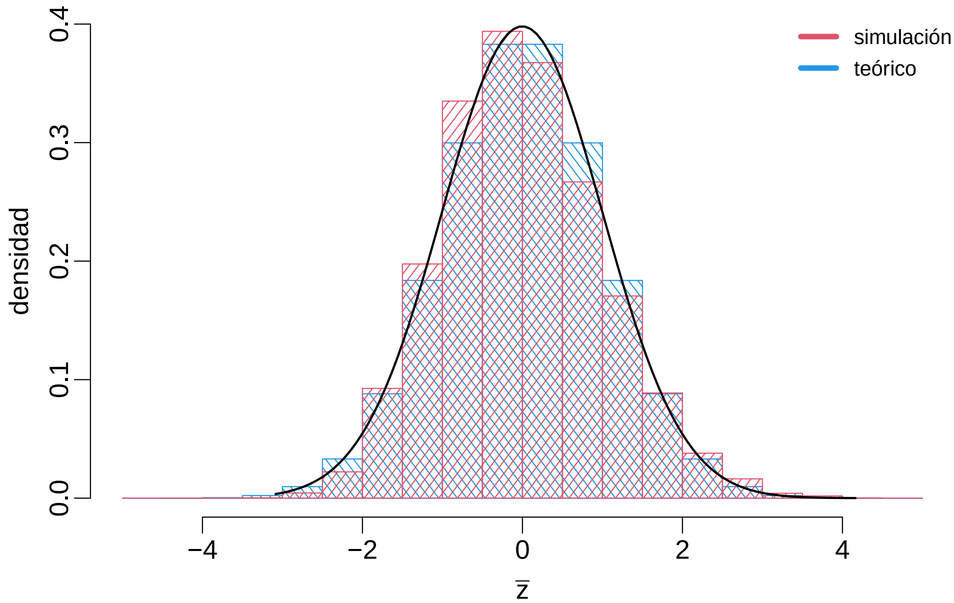

Figura 2: Histograma de promedios estandarizados de muestras de tamaño 25 de una distribución gamma (simulación) sobrepuesto sobre el diagrama de barras de la normal estándar (teórico). Error de aproximación de 0.0808.

III) Calcule los errores de aproximación de la distribución normal teórica (ϕ) a la distribución del promedio de las distintas leyes de probabilidad de la tabla 2 para tamaños de muestra $n = 30, 100, 500$, indicados en la tabla 3.

Tabla 2: Tabla de funciones de densidad para distintas leyes de probabilidad.

Binomial :	$f(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}_{\{0,1,\ldots,n\}}(k)$
Poisson :	$f(n; \lambda) = e^{-\lambda} \frac{\lambda^n}{n!} \mathbb{1}_{\{0,1,2,\ldots\}}(n)$
Normal :	$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathbb{1}_{\mathbb{R}}(x)$
Gamma :	$f(y; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \mathbb{1}_{\mathbb{R}^+}(y)$
Beta :	$f(u; \theta_1, \theta_2) = \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} u^{\theta_1-1} (1-u)^{\theta_2-1} \mathbb{1}_{(0,1)}(u)$

Tabla 3: Error de aproximación de la distribución normal a la del promedio de muestras tamaño n para distintas distribuciones.

distribución	parámetros	$n = 30$	$n = 100$	$n = 500$
Binomial [1]	$p = 0.5$			
$\mathrm{Bin}(25, p)$	$p = 0.7$			
	$p = 0.9$			
Poisson [2]	$\lambda = 1$			
$\mathrm{Po}(\lambda)$	$\lambda = 4$			
	$\lambda = 8$			
Normal [3]				
$\mathrm{N}(\mu, \sigma^2)$	$\mu = 2,\ \sigma^2 = 4$			
Gamma [4]	$\alpha = 1,\ \beta = 3$			
$\mathrm{Gamma}(\alpha, \beta)$	$\alpha = 3,\ \beta = 1$			
	$\alpha = 5,\ \beta = 5$			
Beta [5]	$\theta_1 = 1,\ \theta_2 = 1$			
$\mathrm{Beta}(\theta_1, \theta_2)$	$\theta_1 = 1/2,\ \theta_2 = 2$			
	$\theta_1 = 3,\ \theta_2 = 1/3$			
	$\theta_1 = 1/2,\ \theta_2 = 1/2$			