Teoria

El Teorema del Límite Central nos dice que, si $X_1, X_2, X_3, \ldots, X_n$ son una muestra aleatoria de una población infinita con media μ, varianza σ^2, y generatriz de momentos $M_X(t)$, entonces la distribución de

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

cuando $n \to \infty$, es la distribución normal estándar.

Simulación

Para ilustrar lo visto en la teoría, se realizó una simulación siguiendo los siguientes pasos:

1. Se tomaron $N = 10{,}000$ datos aleatorios de una distribución normal utilizando la función `rnorm`.
2. Se obtiene el promedio de estos datos.
3. Se repiten los pasos 1 y 2 M veces y se hace un histograma de los promedios obtenidos para $M_1 = 500$, $M_2 = 1{,}000$, $M_3 = 5{,}000$ y $M_4 = 10{,}000$.
4. Se repiten los pasos del 1 al 3 para otras tres distribuciones.
5. Comparamos resultados.

Para la simulación del muestreo sobre una distribución normal estándar se utilizó el siguiente script en lenguaje R. Nota: $\mu = 0$ y $\sigma^2 = 1$.

```
# Parámetros de muestreo de datos
N <- 10000
M1 <- 500
M2 <- 1000
M3 <- 5000
M4 <- 10000

# Método de Montecarlo para datos tomados de la distribución
# normal
v1 <- vector()
v2 <- vector()
v3 <- vector()
v4 <- vector()
for(i in 1:M1){
  x <- rnorm(N)
  v1[i] <- (mean(x))/(1/sqrt(N))
```

```
}
for(i in 1:M2){
  x <- rnorm(N)
  v2[i] <- (mean(x))/(1/sqrt(N))
}
for(i in 1:M3){
  x <- rnorm(N)
  v3[i] <- (mean(x))/(1/sqrt(N))
}
for(i in 1:M4){
  x <- rnorm(N)
  v4[i] <- (mean(x))/(1/sqrt(N))
}
# Histogramas para observar y comparar las distribuciones de las
# muestras
par(mfrow = c(2,2), mar=c(4,4,4,4))
hist(v1, main=NULL, xlim=c(-0.04,0.04), xlab="(a)",
     col="darkgreen")
hist(v2, main=NULL, xlim=c(-0.04,0.04), xlab="(b)",
     col="darkgreen")
hist(v3, main=NULL, xlim=c(-0.04,0.04), xlab="(c)",
     col="darkgreen")
hist(v4, main=NULL, xlim=c(-0.04,0.04), xlab="(d)",
     col="darkgreen")
```

El script anterior nos arroja como datos de salida los siguientes histogramas.

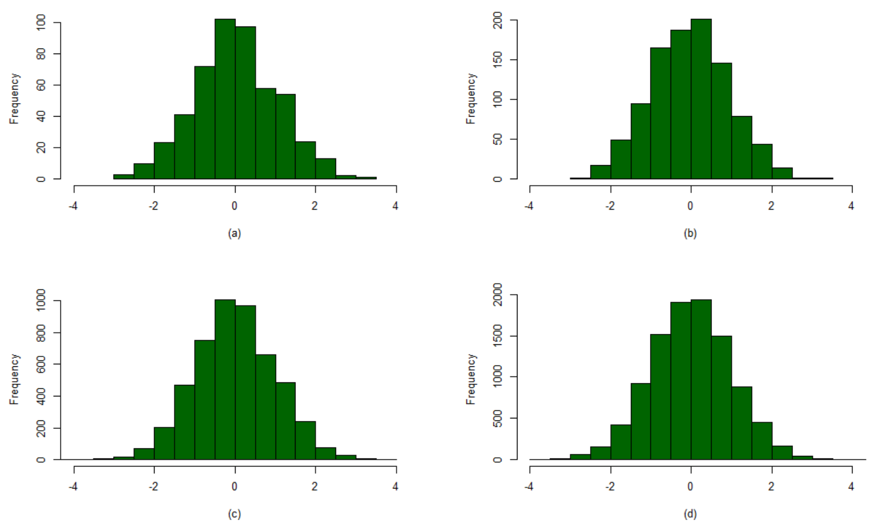

Figura 1: Histogramas para (a) 500 repeticiones, (b) 1,000 repeticiones, (c) 5,000 repeticiones y (d) 10,000 repeticiones del promedio de un muestreo de 1,000 datos de una distribución normal estándar.

Para repetir la simulación con una distribución uniforme, hacemos uso de la función `runif` de R. Los histogramas que nos arroja el script están a continuación. Nota: $\mu = 0.5$ y $\sigma^2 = 1/12$.

```
# Parámetros de muestreo de datos
N <- 10000
M1 <- 500
M2 <- 1000
M3 <- 5000
M4 <- 10000

# Método de Montecarlo para datos tomados de la distribución
# uniforme con alpha = 0 y beta = 1
v1 <- vector()
v2 <- vector()
v3 <- vector()
v4 <- vector()
for(i in 1:M1){
  x <- runif(N)
  v1[i] <- (mean(x)-0.5)/(sqrt(1/12)/sqrt(N))
}
for(i in 1:M2){
  x <- runif(N)
  v2[i] <- (mean(x)-0.5)/(sqrt(1/12)/sqrt(N))
}
for(i in 1:M3){
  x <- runif(N)
  v3[i] <- (mean(x)-0.5)/(sqrt(1/12)/sqrt(N))
}
for(i in 1:M4){
  x <- runif(N)
  v4[i] <- (mean(x)-0.5)/(sqrt(1/12)/sqrt(N))
}
par(mfrow = c(2,2), mar=c(4,4,4,4))
hist(v1, main=NULL, xlim=c(-4,4), xlab="(a)", col="darkgreen")
hist(v2, main=NULL, xlim=c(-4,4), xlab="(b)", col="darkgreen")
hist(v3, main=NULL, xlim=c(-4,4), xlab="(c)", col="darkgreen")
hist(v4, main=NULL, xlim=c(-4,4), xlab="(d)", col="darkgreen")
```

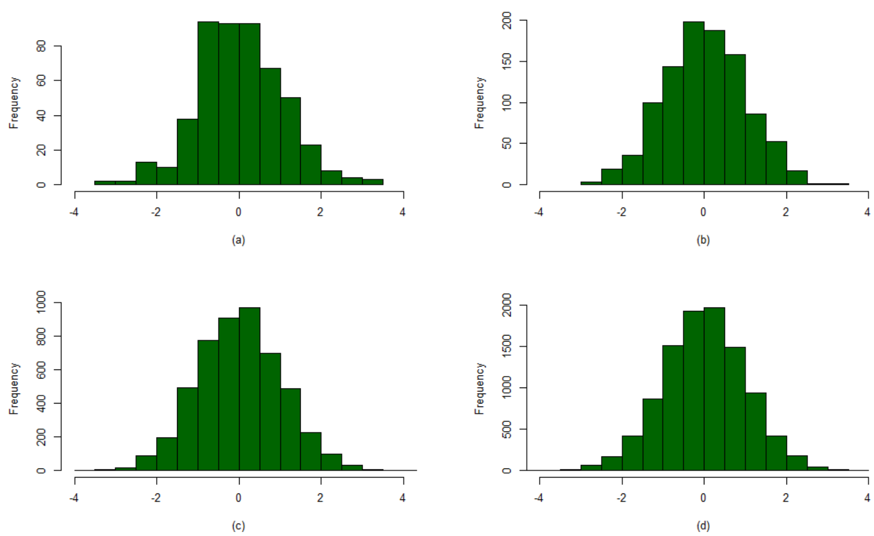

Figura 2: *Histogramas para (a) 500 repeticiones, (b) 1,000 repeticiones, (c) 5,000 repeticiones y (d) 10,000 repeticiones del promedio de un muestreo de 1,000 datos de una distribución uniforme de 0 a 1.*

De la misma forma, hacemos uso de la función `rchisq` para realizar el mismo experimento con la distribución Ji Cuadrada. Los histogramas que nos arroja el script en R son los mostrados a continuación. Nota: $\mu = 3$ y $\sigma^2 = 6$.

```
# Parámetros de muestreo de datos
N <- 10000
M1 <- 500
M2 <- 1000
M3 <- 5000
M4 <- 10000

# Método de Montecarlo para datos tomados de la distribución
# chi cuadrada con nu = 3
v1 <- vector()
v2 <- vector()
v3 <- vector()
v4 <- vector()
for(i in 1:M1){
  x <- rchisq(N, 3)
  v1[i] <- (mean(x)-3)/(sqrt(2*3)/sqrt(N))
}
for(i in 1:M2){
  x <- rchisq(N, 3)
  v2[i] <- (mean(x)-3)/(sqrt(2*3)/sqrt(N))
}
for(i in 1:M3){
  x <- rchisq(N, 3)
  v3[i] <- (mean(x)-3)/(sqrt(2*3)/sqrt(N))
}
for(i in 1:M4){
  x <- rchisq(N, 3)
  v4[i] <- (mean(x)-3)/(sqrt(2*3)/sqrt(N))
}
par(mfrow = c(2,2), mar=c(4,4,4,4))
hist(v1, main=NULL, xlim=c(-4,4), xlab="(a)", col="darkgreen")
hist(v2, main=NULL, xlim=c(-4,4), xlab="(b)", col="darkgreen")
hist(v3, main=NULL, xlim=c(-4,4), xlab="(c)", col="darkgreen")
hist(v4, main=NULL, xlim=c(-4,4), xlab="(d)", col="darkgreen")
```

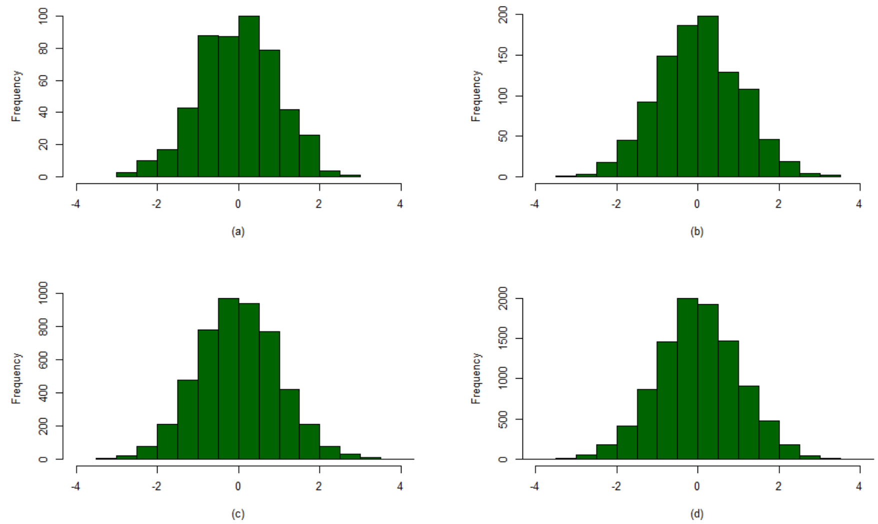

Figura 3: *Histogramas para (a) 500 repeticiones, (b) 1,000 repeticiones, (c) 5,000 repeticiones y (d) 10,000 repeticiones del promedio de un muestreo de 1,000 datos de una distribución ji cuadrada con 3 grados de libertad.*

Por último, realizamos los mismos pasos para la distribución exponencial usando la función `rexp` de R para obtener los siguientes histogramas. Nota: $\mu = 2$ y $\sigma^2 = 4$.

```
# Parámetros de muestreo de datos
N <- 10000
M1 <- 500
M2 <- 1000
M3 <- 5000
M4 <- 10000

# Método de Montecarlo para datos tomados de la distribución
# exponencial con teta = 2
v1 <- vector()
v2 <- vector()
v3 <- vector()
v4 <- vector()
for(i in 1:M1){
  x <- rexp(N, rate=0.5)
  v1[i] <- (mean(x)-2)/(2/sqrt(N))
}
for(i in 1:M2){
  x <- rexp(N, rate=0.5)
  v2[i] <- (mean(x)-2)/(2/sqrt(N))
}
for(i in 1:M3){
  x <- rexp(N, rate=0.5)
  v3[i] <- (mean(x)-2)/(2/sqrt(N))
}
for(i in 1:M4){
  x <- rexp(N, rate=0.5)
```

```
  v4[i] <- (mean(x)-2)/(2/sqrt(N))
}
par(mfrow = c(2,2), mar=c(4,4,4,4))
hist(v1, main=NULL, xlim=c(-4,4), xlab="(a)", col="darkgreen")
hist(v2, main=NULL, xlim=c(-4,4), xlab="(b)", col="darkgreen")
hist(v3, main=NULL, xlim=c(-4,4), xlab="(c)", col="darkgreen")
hist(v4, main=NULL, xlim=c(-4,4), xlab="(d)", col="darkgreen")
```

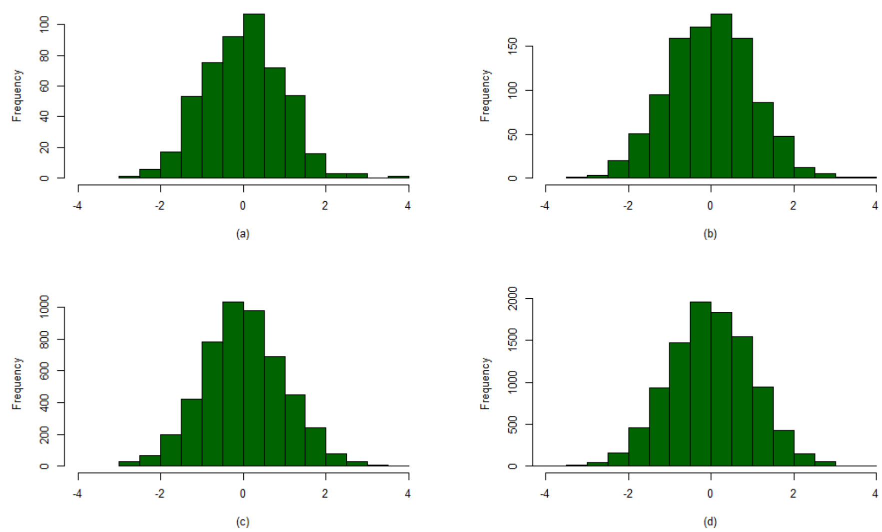

Figura 4: *Histogramas para (a) 500 repeticiones, (b) 1,000 repeticiones, (c) 5,000 repeticiones y (d) 10,000 repeticiones del promedio de un muestreo de 1,000 datos de una distribución exponencial con* $\theta = 2$.

Comentarios

Si bien las simulaciones no son suficientes para demostrar el Teorema del Límite Central, nos ayudan a ver que hay una tendencia a parecerse a la distribución normal estándar cuando M tiende a infinito.