# Tarea 10: Bootstrap Paramétrico

Rodrigo Alan García Pérez

2024-11-12

## Bootstrap paramétrico

Sean $X_1, \ldots, X_n \sim N(\mu, 1)$. Sea $\theta = e^{\mu}$, simula una muestra de $n = 100$ observaciones usando $\mu = 5$.

- Usa el método delta para calcular $ee$ de $\hat{\theta}$ y crea un intervalo del 95% de confianza para $\hat{\theta}$. Pista: $ee(\hat{\mu}) = \frac{1}{\sqrt{n}}$

```r
# Generar la muestra
set.seed(123)
n <- 100
mu <- 5
muestra <- rnorm(n, mean = mu, sd = 1)

# Estimacion de theta
mu_hat <- mean(muestra)
theta_hat <- exp(mu_hat)

# Metodo delta
se_mu_hat <- 1 / sqrt(n) # error estandar de mu gorro
se_theta_hat <- theta_hat * se_mu_hat #g'(mu) * EE(mu) = Error estándar de theta_hat

# Calcular el intervalo de confianza del 95%
ci_delta <- round( c(theta_hat - 1.96 * se_theta_hat, theta_hat + 1.96 * se_theta_hat), 3)
ci_delta
```

```
## [1] 130.614 194.297
```

- Repite el inciso anterior usando bootstrap paramétrico.

```r
set.seed(123)
B <- 1000  # Número de muestras bootstrap
theta_bootstrap_p <- numeric(B) # Inicializamos vector

# Ciclo para generar muestras bootstrap paramétricas
for (i in 1:B){
  muestra_bootstrap <- rnorm(n, mean = mu_hat, sd = 1)
  mu_hat_bootstrap <- mean(muestra_bootstrap)
  theta_bootstrap_p[i] <- exp(mu_hat_bootstrap)
}
```

```r
# Intervalo de confianza
ci_bootstrap_param <- quantile(theta_bootstrap_p, probs = c(0.025, 0.975))
ci_bootstrap_param
```

```
##     2.5%    97.5%
## 134.5653 194.4259
```

- Finalmente usa bootstrap no paramétrico.

```r
B <- 1000  # Número de muestras bootstrap
theta_bootstrap_np <- numeric(B) # Inicializamos vector

# Ciclo para generar muestras bootstrap no paramétricas
for (i in 1:B){
  muestra_bootstrap <- sample(muestra, size = n, replace = TRUE) # Muestra con reemplazo de datos origi
  mu_hat_bootstrap <- mean(muestra_bootstrap)
  theta_bootstrap_np[i] <- exp(mu_hat_bootstrap)
}

#Intervalo de confianza
# Paso 3: Calcular el intervalo de confianza del 95% usando percentiles bootstrap
ci_bootstrap_np <- quantile(theta_bootstrap_np, probs = c(0.025, 0.975))
ci_bootstrap_np
```

```
##     2.5%    97.5%
## 135.4672 194.7000
```

- Compara tus respuestas.

```r
library(tidyverse)

tabla_intervalos <- tibble(
  Metodo =  c("Bootstrap No Paramétrico", "Bootstrap Paramétrico", "Método Delta"),
  Limite_inferior = format(c(ci_bootstrap_np[1],
                             ci_bootstrap_param[1],
                             ci_delta[1]),
                           nsmall = 3),
  Limite_superior = format(c(ci_bootstrap_np[2],
                             ci_bootstrap_param[2],
                             ci_delta[2]),
                           nsmall = 3)
) %>%
  mutate( Intervalo = format(as.numeric(Limite_superior) - as.numeric(Limite_inferior), nsmall = 3))

tabla_intervalos
```

```
## # A tibble: 3 x 4
##   Metodo                   Limite_inferior Limite_superior Intervalo
##   <chr>                    <chr>           <chr>           <chr>
## 1 Bootstrap No Paramétrico 135.4672        194.7000        59.2328
## 2 Bootstrap Paramétrico    134.5653        194.4259        59.8606
## 3 Método Delta             130.6140        194.2970        63.6830
```

Se puede observar que de los tres métodos el método Delta es el que tiene el intervalo de confianza más amplio. Esto sugiere mayor incertidumbre en la estimación. El mejor método fue el bootstrap no paramétrico sin embargo se comporto de manera muy similar al paramétrico.

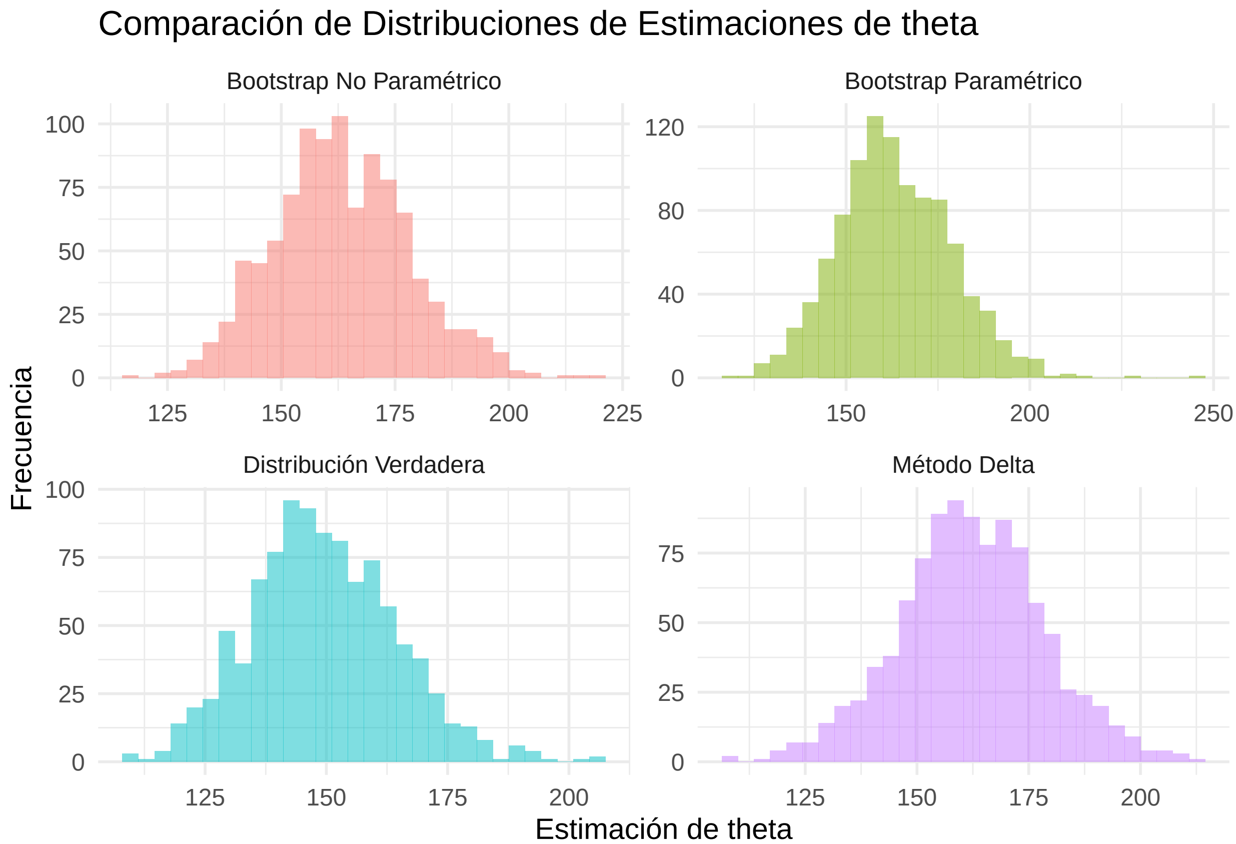
- Realiza un histograma de replicaciones bootstrap para cada método, estas son estimaciones de la distribución de $\hat{\theta}$. El método delta también nos da una aproximación a esta distribución: Normal$(\hat{\theta}, \hat{ee}^2)$. Compáralos con la verdadera distribución de $\hat{\theta}$ (que puedes obtener vía simulación). ¿Cuál es la aproximación más cercana a la verdadera distribución?

```r
# Generar valores de theta usando metodo delta
theta_delta <- rnorm(B, mean = theta_hat, sd = se_theta_hat)

# Simulacion de theta
theta_simulacion <- numeric(B)
for (i in 1:B){
  muestra_sim <- rnorm(n, mean = mu, sd = 1)
  mu_hat_sim <- mean(muestra_sim)
  theta_simulacion[i] <- exp(mu_hat_sim)
}

data <- data.frame(
  theta = c(theta_bootstrap_np,
            theta_bootstrap_p,
            theta_delta,
            theta_simulacion),
  metodo = factor(rep(c("Bootstrap No Paramétrico",
                        "Bootstrap Paramétrico",
                        "Método Delta",
                        "Distribución Verdadera"),
                      each = B))
)

# Graficar histogramas para cada método
ggplot(data, aes(x = theta, fill = metodo)) +
  geom_histogram(bins = 30, alpha = 0.5, position = "identity") +
  facet_wrap(~ metodo, scales = "free") +
  labs(title = "Comparación de Distribuciones de Estimaciones de theta",
       x = "Estimación de theta",
       y = "Frecuencia") +
  theme_minimal() +
  theme(legend.position = "none")
```




Tomando la distribución verdadera como referencia, se puede observar que el bootstrap no paramétrico es la mejor aproximación de la distribución verdadera de $\theta$ su forma es la más cercana a la distribución verdadera, capturando tanto la simetría como las colas de la distribución.