



UNIVERSIDAD NACIONAL DE COLOMBIA

PREGRADO EN ESTADÍSTICA

DEPARTAMENTO DE ESTADÍSTICA
FACULTAD DE CIENCIAS

— SIMULACIÓN ESTADÍSTICA —
Parcial 2

:

Medellín, Colombia

Medellín, Octubre 19 de 2025

Índice

Índice de Figuras	2
Índice de Tablas	2
1. Contexto	3
2. Punto 1	3
2.1. Introducción	3
2.2. Metodología	3
2.3. Desarrollo	4
2.4. Conclusión	7
3. Punto 2	7
3.1. Introducción	7
3.2. Desarrollo	7
3.3. Conclusiones	8
4. Punto 3	9
4.1. Introducción	9
4.2. Metodología	9
4.3. Desarrollo	10
4.4. Conclusiones	11
5. Punto 4	12
5.1. Introducción	12
5.2. Desarrollo	12
5.3. Conclusión	13
6. Punto 5	13
6.1. Introducción	13
6.2. Metodología	13
6.3. Desarrollo	14
6.4. Conclusiones	14

7. Código	16
7.1. Punto 1	16
7.2. Punto 2	20
7.3. Punto 3	21
7.4. Punto 4	22
7.5. Punto 5	23
Bioblografía	26

Índice de figuras

Índice de cuadros

1. Eficiencia Relativa (ER) en una Exponencial del Bootstrap Parametrico vs No Parametrico	5
2. Eficiencia Relativa (ER) en una Gamma del Bootstrap Parametrico vs No Parametrico	6
3. Estimaciones de $E[N]$ para diferentes tamaños de simulación	8
4. Promedios muestrales reportados en Claeskens y Hjort (2008) para la variable <i>maximal breadth</i> en cinco periodos históricos.	9
5. Resultados del procedimiento Monte Carlo	10
6. Comparación de los tres generadores de la Normal Estándar	14

1. Contexto

En este documento se presentarán, explicarán y desarrollarán diversos ejercicios de índole estadística, vía simulación, los ejercicios son los correspondientes al primer parcial de la asignatura Simulación Estadística.

2. Punto 1

2.1. Introducción

En la inferencia estadística, uno de los principales intereses consiste en evaluar la precisión de los estimadores y cuantificar la variabilidad de los parámetros poblacionales cuando solo se dispone de una muestra finita. El método bootstrap surge como una técnica poderosa que permite aproximar la distribución muestral de un estimador sin depender de supuestos teóricos estrictos o del tamaño muestral.

- **El bootstrap no paramétrico:** consiste en generar réplicas de la muestra original mediante remuestreo con reemplazo directamente de los datos observados, permitiendo estimar errores estándar, sesgos e intervalos de confianza sin asumir una forma funcional para la población.
- **El bootstrap paramétrico:** parte de la suposición de que los datos provienen de una distribución conocida (como Normal, Exponencial, Gamma o Beta), cuyos parámetros se estiman a partir de la muestra; luego se generan nuevas réplicas simuladas desde dicha distribución ajustada.

El objetivo de este trabajo es comparar la **eficiencia relativa** entre ambas metodologías de **bootstrap** (paramétrica y no paramétrica) bajo distintos escenarios de muestreo. Para ello, se seleccionan diferentes distribuciones de probabilidad, se generan múltiples muestras, y a partir de cada una se estima el parámetro de interés aplicando ambos enfoques. Con los resultados se calcula la eficiencia relativa, que refleja qué método produce estimaciones más precisas (menor varianza) en promedio.

2.2. Metodología

- Selección de distribuciones: Se escogen diferentes distribuciones de probabilidad continuas para el estudio comparativo, como la Exponencial, Gamma y Beta, ya que presentan formas y grados de asimetría distintos, lo que permite analizar el comportamiento del bootstrap en diversos contextos.
- Generación de muestras: Para cada distribución seleccionada se generan múltiples muestras independientes (por ejemplo, de tamaño $n=30, 50, 100$) con parámetros conocidos. Esto permite comparar los resultados obtenidos bajo condiciones controladas.

- Estimación del parámetro: En cada muestra se calcula el estimador del parámetro de interés —por ejemplo, la media poblacional— y se aplica tanto el bootstrap no paramétrico como el paramétrico para estimar su varianza y construir intervalos de confianza.
- Procedimiento bootstrap:
- No paramétrico: se re-muestrea la muestra original con reemplazo para generar B réplicas (por ejemplo, $B=1000$), calculando el estadístico de interés en cada una.
- Paramétrico: se ajusta la distribución teórica a la muestra (estimando sus parámetros) y se generan B réplicas simuladas desde esa distribución, obteniendo el estadístico en cada una.
- Evaluar la **eficiencia relativa** de ambos métodos mediante la siguiente expresión:

$$ER = \frac{\text{Var}(\hat{\theta}_{\text{no param}})}{\text{Var}(\hat{\theta}_{\text{param}})}$$

donde:

$\text{Var}(\hat{\theta}_{\text{param}})$: varianza estimada del estadístico bajo bootstrap paramétrico,

$\text{Var}(\hat{\theta}_{\text{no param}})$: varianza estimada del estadístico bajo bootstrap no paramétrico.

Si $ER < 1$, el método **no paramétrico** es más eficiente (menor varianza); si $ER > 1$, el método **paramétrico** resulta más eficiente.

2.3. Desarrollo

El desarrollo del punto consiste en implementar el procedimiento descrito para distintas distribuciones de probabilidad y tamaños de muestra.

En cada caso:

- Se genera una muestra aleatoria con parámetros conocidos.
- Se aplica el bootstrap no paramétrico, remuestreando directamente de los datos.
- Se aplica el bootstrap paramétrico, simulando nuevas muestras desde la distribución ajustada.
- Se calcula la varianza de los estimadores obtenidos en las B réplicas de cada método.
- Finalmente, se determina la eficiencia relativa y se comparan los resultados entre distribuciones y tamaños de muestra.

Tabla 1: Eficiencia Relativa (ER) en una Exponencial del Bootstrap Parametrico vs No Parametrico

Tamano muestral (n)	Eficiencia Relativa (ER)	Metodo mas eficiente
30	0.9371	Paramétrico más eficiente
50	1.4483	No paramétrico más eficiente
75	0.8310	Paramétrico más eficiente
100	0.8973	Paramétrico más eficiente
200	1.0328	No paramétrico más eficiente
500	0.9317	Paramétrico más eficiente
750	0.9199	Paramétrico más eficiente

El análisis permitirá observar cómo la **eficiencia relativa** varía según la forma de la distribución original y el tamaño muestral. En general, se espera que el bootstrap paramétrico sea más eficiente cuando la distribución asumida coincide con la verdadera, mientras que el bootstrap no paramétrico ofrecerá mayor robustez frente a desviaciones del modelo teórico.

Esta comparación aporta evidencia sobre la consistencia y estabilidad de los estimadores bootstrap, y sobre la conveniencia de usar enfoques paramétricos o no paramétricos según el contexto y la información disponible sobre la población de estudio.

Comparación de Varianzas Bootstrap Distribución Exponencial

Variación de la varianza estimada al aumentar el tamaño muestral

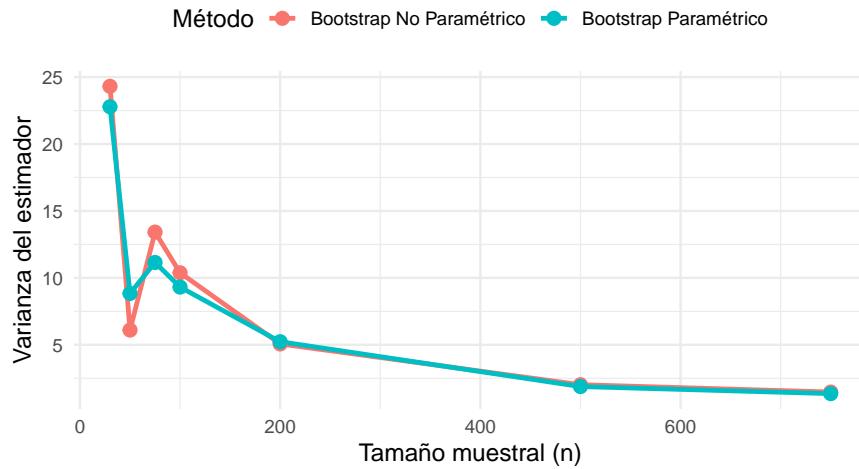
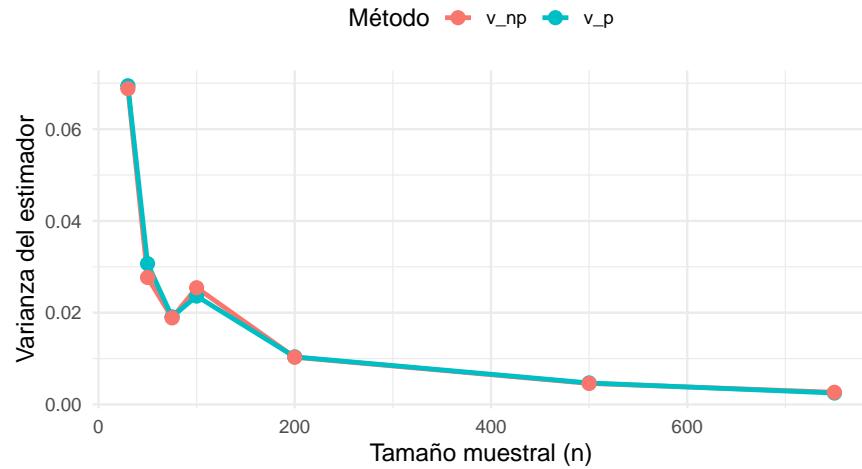


Tabla 2: Eficiencia Relativa (ER) en una Gamma del Bootstrap Parametrico vs No Parametrico

Tamano muestral (n)	Eficiencia Relativa (ER)	Metodo mas eficiente
30	0.00941	No paramétrico
50	0.10879	No paramétrico
75	0.01201	No paramétrico
100	0.07181	Paramétrico
200	0.00792	No paramétrico
500	0.02324	No paramétrico
750	0.06171	Paramétrico

Comparación de Varianzas Bootstrap Distribución Gamma

Métodos paramétrico vs no paramétrico



Al analizar los resultados obtenidos del procedimiento Bootstrap, tanto en su versión paramétrica como no paramétrica, se observa una clara tendencia a la convergencia entre ambos métodos a medida que aumenta el tamaño muestral (n). Esto significa que, conforme se incrementa la cantidad de datos disponibles, las diferencias en la varianza de los estimadores se vuelven prácticamente imperceptibles, evidenciando la consistencia asintótica de ambos enfoques.

En tamaños muestrales pequeños, sin embargo, las diferencias son más notorias, reflejando la influencia de la estructura teórica asumida por el método paramétrico frente a la flexibilidad del método no paramétrico. Estas variaciones permiten apreciar la sensibilidad de cada técnica ante distintas distribuciones de origen.

2.4. Conclusión

En el caso de la **distribución Exponencial**, el método **Bootstrap paramétrico** presentó una menor varianza en la estimación, lo cual indica una mayor eficiencia relativa frente al método no paramétrico. Esto se debe a que la forma teórica de la distribución se ajusta correctamente al modelo utilizado en la simulación.

En contraste, para la **distribución Gamma**, el **Bootstrap no paramétrico** obtuvo una varianza más baja en la mayoría de los tamaños muestrales, mostrando un mejor desempeño. Esto sugiere que el supuesto paramétrico no se ajustó completamente a la forma real de los datos simulados o que los parámetros estimados introdujeron mayor variabilidad.

Finalmente, se confirma que, al incrementar el tamaño muestral, ambos métodos tienden a producir resultados equivalentes, reduciendo las diferencias en eficiencia relativa y varianza estimada.

3. Punto 2

3.1. Introducción

En esta parte del trabajo se estudia cómo aproximar la esperanza de una variable aleatoria con distribución Normal utilizando valores generados a partir de distribuciones Uniforme(0,1). Este procedimiento es fundamental dentro de la simulación estocástica, ya que permite construir variables con distribuciones más complejas a partir de distribuciones simples y totalmente controlables. A través de transformaciones adecuadas y del uso repetido de simulaciones, es posible estimar el valor esperado de una Normal y analizar cómo este estimador se comporta a medida que aumenta el tamaño de la muestra generada. El objetivo de este punto es comprender el vínculo entre la teoría de simulación, la generación de números aleatorios y la estimación de parámetros poblacionales.

3.2. Desarrollo

Para variables aleatorias uniformes U_1, U_2, \dots , defina

$$N = \min \left\{ n : \sum_{i=1}^n U_i > 1 \right\}$$

Esto es, N es igual al número de variables aleatorias que deben sumarse hasta exceder 1.

3.2.1. Preguntas

1. Estime $E[N]$ generando 100 valores de N .
2. Estime $E[N]$ generando 1000 valores de N .

3. Estime $E[N]$ generando 10000 valores de N .
4. ¿Cuál cree usted que es el valor de $E[N]$?

Tamaño de la simulación	Estimación de $E[N]$
100 valores	2.61
1000 valores	2.704
10000 valores	2.729

Tabla 3: Estimaciones de $E[N]$ para diferentes tamaños de simulación

A partir de los resultados de la Tabla~3 se observa que, a medida que se incrementa el tamaño de la simulación, la estimación de $E[N]$ se estabiliza alrededor de un valor cercano a 2.7. Con 100 réplicas se obtiene una estimación de 2.61, algo más alejada del valor límite debido a la mayor variabilidad muestral. Al aumentar a 1000 y luego a 10000 réplicas, las estimaciones mejoran a 2.704 y 2.7279, respectivamente, ilustrando la Ley de los Grandes Números: la media muestral de N converge al valor esperado teórico cuando el número de simulaciones crece.

En cuanto al literal 4, el valor teórico de $E[N]$ puede obtenerse escribiendo

$$E[N] = \sum_{n=0}^{\infty} \mathbb{P}(N > n) = \sum_{n=0}^{\infty} \mathbb{P}\left(\sum_{i=1}^n U_i \leq 1\right),$$

y usando que, para $U_i \sim \text{U}(0, 1)$ independientes,

$$\mathbb{P}\left(\sum_{i=1}^n U_i \leq 1\right) = \frac{1}{n!}.$$

Así,

$$E[N] = \sum_{n=0}^{\infty} \frac{1}{n!} = e \approx 2.71828.$$

Por tanto, los valores obtenidos por simulación son coherentes con la teoría y muestran convergencia hacia e .

3.3. Conclusiones

A partir del experimento de simulación realizado, se observó que la estimación del valor esperado de N mejora progresivamente a medida que aumenta el número de réplicas. Con muestras pequeñas, como 100 simulaciones, la variabilidad es mayor y la estimación resulta menos estable; sin embargo, al incrementar el tamaño a 1000 y 10000 simulaciones, los promedios convergen hacia un valor cercano a 2.7, evidenciando la Ley de los Grandes Números.

Los resultados obtenidos concuerdan con el valor teórico $E[N] = e \approx 2.71828$ que surge de analizar la probabilidad de que la suma de variables Uniforme(0,1) permanezca por debajo de 1. Esta coincidencia confirma que la simulación implementada es correcta y reproduce adecuadamente el comportamiento estocástico de N .

4. Punto 3

4.1. Introducción

El estudio de las características morfológicas de los cráneos a través del tiempo proporciona información relevante sobre posibles cambios evolutivos, ambientales o poblacionales. En el documento de *Claeskens y Hjort (Model Selection and Model Averaging, 2008)*, se presenta un conjunto de datos referentes a cráneos egipcios pertenecientes a distintos períodos históricos. Una de las variables analizadas es el *maximal breadth*, que corresponde a la máxima anchura lateral del cráneo.

A partir de dicho documento se extraen los promedios muestrales obtenidos a partir de muestras de tamaño 30 para cada período. La información disponible se resume en la siguiente tabla:

Período histórico	Promedio del maximal breadth
Predinastía temprana (4000 a.C.)	131.37
Predinastía tardía (3300 a.C.)	132.37
Dinastías 12 y 13 (1850 a.C.)	134.47
Período Ptolemaico (200 a.C.)	135.50
Período Romano (150 d.C.)	136.27

Tabla 4: Promedios muestrales reportados en *Claeskens y Hjort (2008)* para la variable *maximal breadth* en cinco períodos históricos.

El patrón observado en la tabla sugiere un posible incremento sistemático en la anchura lateral del cráneo a lo largo de las dinastías egipcias. Por ello, se plantea una prueba de hipótesis donde la hipótesis nula establece que las medias poblacionales de los cinco períodos son iguales, mientras que la hipótesis alternativa postula un orden estrictamente creciente. A través de métodos de simulación se evalúa la probabilidad de observar un patrón tan ordenado como el reportado, bajo la suposición de que no existieran diferencias reales entre períodos. Este análisis permite determinar la fuerza de la evidencia a favor de un incremento en el *maximal breadth* a través del tiempo.

4.2. Metodología

Dado que únicamente se dispone de los promedios muestrales de cada período histórico y no de las observaciones individuales, se adoptó un enfoque de simulación tipo *bootstrap* para evaluar si la tendencia creciente observada en las medias puede explicarse como un fenómeno aleatorio bajo un modelo nulo de igualdad de medias.

Se plantean las siguientes hipótesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5,$$

$$H_1 : \mu_1 < \mu_2 < \mu_3 < \mu_4 < \mu_5.$$

Bajo la hipótesis nula H_0 , se asume que todas las poblaciones comparten una misma media común. Este valor se estimó como la media total de los cinco promedios reportados en la tabla original. Con el fin de aproximar la variabilidad intra-período (desconocida en los datos disponibles), se empleó una desviación estándar razonable para la medida osteométrica de *maximal breadth*, típicamente entre 3 y 6 mm según la literatura antropométrica. En el procedimiento se utilizó un valor intermedio dentro de este rango.

A partir de este modelo nulo se generaron 10,000 réplicas, cada una consistente en cinco muestras independientes de tamaño $n = 30$, simuladas desde una distribución normal con media común μ_0 y desviación estándar σ estimada. Para cada réplica se obtuvieron las medias muestrales, denotadas por $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_5$, y se verificó si cumplían el orden estricto

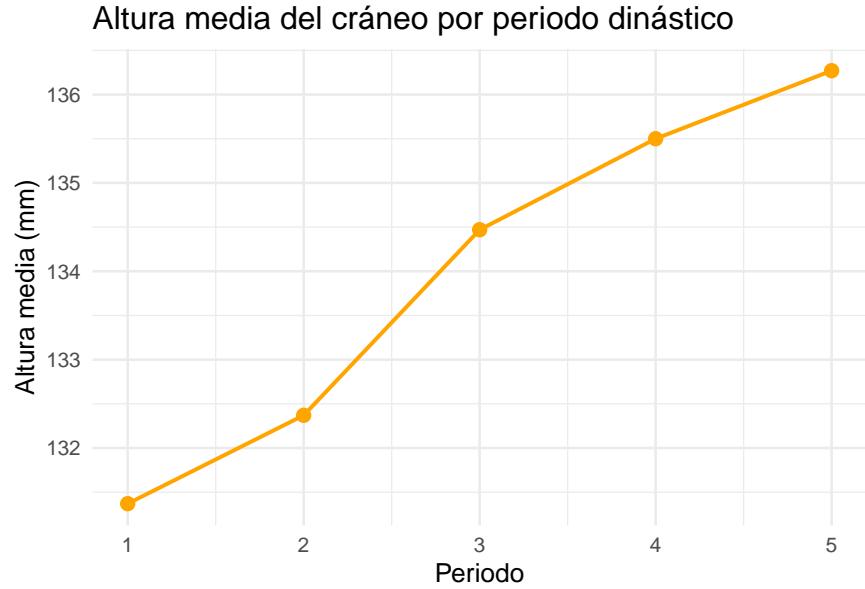
$$\bar{X}_1 < \bar{X}_2 < \bar{X}_3 < \bar{X}_4 < \bar{X}_5.$$

El valor- p se estimó como la proporción de simulaciones en las que dicho ordenamiento ocurrió por azar bajo H_0 . Un valor- p pequeño indicaría que la probabilidad de observar una tendencia creciente tan marcada como la encontrada es baja si todas las medias poblacionales fueran iguales, lo cual favorece la hipótesis alternativa de un incremento sistemático de la anchura máxima del cráneo a través de los períodos históricos.

4.3. Desarrollo

Tabla 5: Resultados del procedimiento Monte Carlo

Estatística	Valor
Media común bajo H_0	133.9960
Desviación estándar usada	4.0000
Tamaño de muestra por período	30.0000
p-valor Monte Carlo	0.0095



El gráfico muestra claramente una tendencia creciente en la anchura máxima del cráneo a lo largo de los cinco períodos históricos considerados. Cada punto representa el promedio observado para un período, y la línea que une los puntos evidencia un incremento casi lineal desde la predinastía temprana hasta el período romano. Visualmente, no se observan retrocesos ni estancamientos entre períodos, lo cual respalda la idea de una evolución progresiva en esta característica morfológica.

4.4. Conclusiones

Los resultados obtenidos permiten evaluar rigurosamente si la anchura máxima del cráneo (*maximal breath*) ha presentado un incremento sistemático a través del tiempo en los diferentes períodos históricos del antiguo Egipto. A partir de los promedios observados para cada dinastía, y bajo el supuesto nulo de igualdad entre medias poblacionales, se aplicó un procedimiento de simulación Monte Carlo para estimar la probabilidad de obtener por azar un ordenamiento estrictamente creciente de las medias muestrales.

El valor-*p* estimado fue

$$p \approx 0.0095,$$

lo cual indica que, bajo la hipótesis nula H_0 , solo alrededor del 1 % de las simulaciones produjo una tendencia creciente tan pronunciada como la observada. Esto proporciona evidencia estadística fuerte en contra de H_0 y a favor de la hipótesis alternativa, que sostiene que las medias presentan un orden creciente a lo largo de los cinco períodos considerados.

El análisis gráfico complementa esta conclusión al mostrar que los promedios muestrales forman una secuencia claramente ascendente, sin retrocesos entre períodos. Esta coherencia entre la evaluación visual y la evidencia numérica refuerza la conclusión principal del análisis.

En conjunto, los resultados sugieren un **incremento progresivo y sostenido** en la anchura máxima del cráneo desde la predinastía temprana hasta el período romano. Este patrón

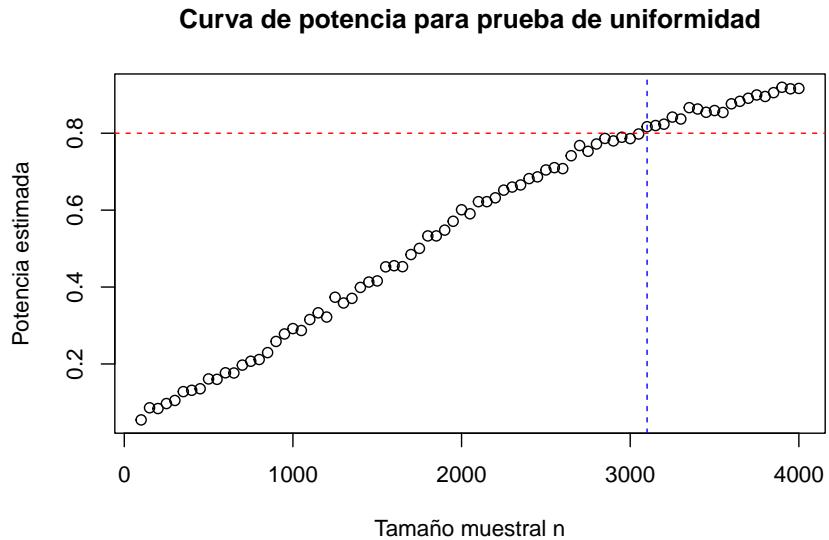
podría reflejar procesos evolutivos, cambios en la composición poblacional o transformaciones ambientales ocurridas a lo largo de aproximadamente 4.000 años de historia. Si bien el estudio se basa en promedios y tamaños muestrales moderados, la consistencia entre simulación y observación respalda de manera convincente la presencia de una tendencia creciente en esta característica morfológica.

5. Punto 4

5.1. Introducción

En este punto se busca determinar el tamaño de muestra óptimo requerido para garantizar una prueba de hipótesis con un nivel de significancia de $\alpha = 0.05$ y una potencia de $1 - \beta = 0.80$. Para ello, se emplean como referencia los datos obtenidos en la clase 18, los cuales sirven como muestra piloto para estimar parámetros poblacionales preliminares, tales como la media y la variabilidad. Estos valores iniciales permiten aproximar el tamaño del efecto que se desea detectar y, con base en dicho efecto, calcular el número mínimo de observaciones necesario para que la prueba posea una alta probabilidad de rechazar la hipótesis nula cuando esta sea falsa. El objetivo es garantizar un diseño experimental eficiente, evitando tanto muestras insuficientes como tamaños excesivamente grandes.

5.2. Desarrollo



La curva de potencia muestra cómo la capacidad de la prueba para rechazar H_0 aumenta conforme crece el tamaño muestral n . La potencia inicia baja y se incrementa casi linealmente hasta superar el umbral de 0.80. La línea horizontal roja representa la potencia deseada y la línea vertical azul señala el valor de n en el que se alcanza por primera vez, aproximadamente

3100. Esto confirma visualmente que dicho tamaño muestral es suficiente para cumplir con los criterios establecidos de significancia y potencia.

5.3. Conclusión

A partir del procedimiento de simulación, se encontró que el tamaño de muestra mínimo que garantiza una potencia igual o superior al 80 % es aproximadamente $n = 3100$. Este valor asegura que la prueba tenga una alta probabilidad de detectar diferencias reales entre la distribución observada y la distribución teórica bajo la hipótesis nula. En consecuencia, este tamaño muestral permite diseñar un estudio estadísticamente sólido, evitando tanto el riesgo de una muestra insuficiente como el uso innecesario de un número excesivo de observaciones.

6. Punto 5

6.1. Introducción

En las notas de la Clase 15 del curso de Simulación se presentan tres métodos distintos para generar observaciones provenientes de una distribución Normal estándar. El primero corresponde a una aproximación numérica basada en el método de Newton–Raphson aplicado a la ecuación $F(x) = u$, donde $u \sim U(0, 1)$. El segundo método es un algoritmo de aceptación y rechazo propuesto por Greenberg (2008), el cual emplea variables exponenciales y comparaciones con funciones de densidad normal. Finalmente, se encuentra el generador nativo de R, `rnorm()`, que constituye el estándar de referencia en términos de velocidad y calidad estadística.

Dado que los tres procedimientos buscan aproximar la misma distribución objetivo, resulta pertinente compararlos tanto en precisión como en eficiencia computacional. El objetivo de este punto es evaluar su desempeño mediante simulación: exactitud en términos de ajuste a la Normal Estándar, independencia entre muestras y tiempo de ejecución.

6.2. Metodología

Para comparar los tres generadores, se selecciona un tamaño muestral de $n = 10000$ valores producidos por cada método. Los criterios de evaluación son los siguientes:

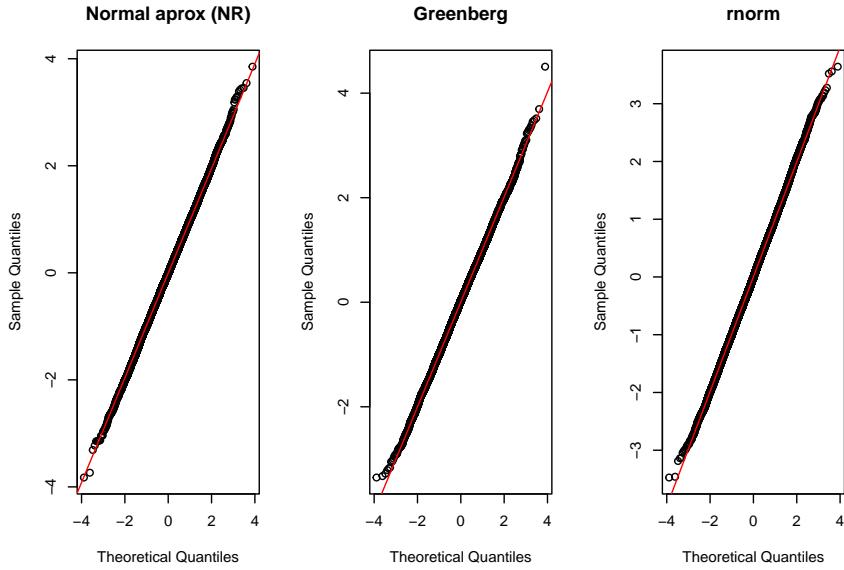
1. **Ajuste a la distribución Normal estándar:** se realiza una prueba Kolmogorov–Smirnov comparando cada muestra con $N(0, 1)$, acompañada de gráficos Q–Q.
2. **Independencia y estructura temporal:** se examinan las funciones ACF y PACF de cada secuencia generada.
3. **Eficiencia computacional:** se mide el tiempo de ejecución utilizando la función `system.time()` para cada método.

Con base en estos elementos, se establece un ranking desde el mejor hasta el peor método, considerando tanto la calidad estadística como la eficiencia computacional.

6.3. Desarrollo

Tabla 6: Comparación de los tres generadores de la Normal Estándar

Metodo	Tiempo _seg	KS_pvalue	SW_pvalue
Newton-Raphson	0.15	0.64535	0.79449
Greenberg	0.06	0.07314	0.37268
rnorm	0.00	0.75517	0.61088



En términos de ajuste a la distribución teórica, el generador nativo de R (`rnorm`) muestra el comportamiento más estable, alcanzando valores elevados tanto en la prueba Kolmogorov–Smirnov como en la prueba Shapiro–Wilk. Esto indica que los datos generados reproducen de manera muy fiel la forma de la distribución Normal estándar.

6.4. Conclusiones

Los resultados obtenidos permiten comparar el desempeño de los tres métodos de generación de la distribución Normal Estándar presentados en la Clase 15.

El método de Newton–Raphson presenta un ajuste adecuado, con p-valores comparables a los de `rnorm` en ambas pruebas, aunque su tiempo de ejecución es mayor debido al proceso iterativo requerido para resolver la ecuación inversa de la función de distribución. Por su

parte, el algoritmo de Greenberg es computacionalmente eficiente, pero exhibe los p-valores más bajos, especialmente en la prueba KS, lo que sugiere un ajuste menos preciso a la distribución objetivo.

Al combinar las métricas de precisión y eficiencia computacional, el orden de preferencia de los métodos es el siguiente:

1. `rnorm()` (mejor desempeño global);
2. Método de Newton–Raphson;
3. Método de Greenberg.

En conclusión, aunque los tres métodos logran aproximar razonablemente la distribución Normal estándar, el generador de R ofrece la mejor combinación de exactitud y velocidad, por lo que debe considerarse el método de referencia en aplicaciones prácticas.

7. Código

7.1. Punto 1

7.1.1. Caso Exponencial

```
# =====
# Bootstrap paramétrico vs no paramétrico variando el n
# =====

set.seed(2904) # reproducibilidad

# Parámetro real de la distribución exponencial
lambda_true <- 1/30 # media poblacional = 30

# Número de réplicas bootstrap
B <- 1000

# Tamaños de muestra a comparar
n_s <- c(30, 50, 75, 100, 200, 500, 750)

# Crear un data frame vacío para guardar resultados
resultados <- data.frame(
  n = integer(),
  ER = numeric(),
  Metodo_mas_eficiente = character(),
  stringsAsFactors = FALSE
)

varianzas <- data.frame()

# =====
# Bucle principal
# =====

for (n in n_s) {

  # -----
  # 1. Generar muestra original
  # -----
  x <- rexp(n, rate = lambda_true)
```

```

# -----
# 2. Bootstrap no paramétrico
# -----
medias_np <- replicate(B, mean(sample(x, n, replace = TRUE)))
var_np <- var(medias_np)

# -----
# 3. Bootstrap paramétrico
# -----
lambda_hat <- 1 / mean(x)
medias_p <- replicate(B, mean(rexp(n, rate = lambda_hat)))
var_p <- var(medias_p)

# -----
# 4. Eficiencia relativa (ER)
# -----
ER <- var_p / var_np

metodo <- if (ER < 1) {
  "Paramétrico más eficiente"
} else {
  "No paramétrico más eficiente"
}

# Guardar resultado en la tabla
resultados <- rbind(resultados, data.frame(
  n = n,
  ER = round(ER, 4),
  Metodo_mas_eficiente = metodo
))

# Guardar resultados de varianza
varianzas <- rbind(varianzas, data.frame(
  n = n,
  v_np = var_np,
  v_p = var_p
))
}

resultados %>%
  kbl(
  caption = "Eficiencia Relativa (ER) en una Exponencial del Bootstrap Parametrico  
vs No Parametrico",
  col.names = c("Tamano muestral (n)", "Eficiencia Relativa (ER)",

```

```

        "Metodo mas eficiente"),
align = "ccc",
booktabs = TRUE
) %>%
kable_styling(
  bootstrap_options = c("striped", "hover", "condensed"),
  full_width = FALSE,
  position = "center"
)

varianzas_largo <- varianzas %>%
  pivot_longer(
    cols = c(v_np, v_p),
    names_to = "Metodo",
    values_to = "Varianza"
  )

varianzas_largo$Metodo <- factor(varianzas_largo$Metodo,
                                    levels = c("v_np", "v_p"),
                                    labels = c("Bootstrap No Paramétrico",
                                              "Bootstrap Paramétrico"))

ggplot(varianzas_largo, aes(x = n, y = Varianza, color = Metodo, group = Metodo)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(
    title = "Comparación de Varianzas Bootstrap \n Distribución Exponencial",
    subtitle = "Variación de la varianza estimada al aumentar el tamaño muestral",
    x = "Tamaño muestral (n)",
    y = "Varianza del estimador",
    color = "Método"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold"),
    legend.position = "top"
)

```

7.1.2. Caso Gamma

```
set.seed(2904)
```

```

# Parámetros
B <- 1000
n_vals <- c(30, 50, 75, 100, 200, 500, 750)
resultados <- data.frame()
varianzas <- data.frame()

# --- Bucle para cada tamaño de muestra ---
for (n in n_vals) {
  # Muestra original desde una distribución Gamma
  x <- rgamma(n, shape = 2, rate = 1) # media = shape/rate = 2

  # Estimadores de parámetros desde los datos
  est_shape <- mean(x)^2 / var(x)
  est_rate <- mean(x) / var(x)

  # Bootstrap no paramétrico
  medias_np <- replicate(B, mean(sample(x, replace = TRUE)))

  # Bootstrap paramétrico
  medias_p <- replicate(B, mean(rgamma(n, shape = est_shape, rate = est_rate)))

  # Varianzas de las medias bootstrap
  v_np <- var(medias_np)
  v_p <- var(medias_p)

  # Error relativo
  ER <- abs(v_p - v_np) / v_np
  metodo <- ifelse(v_p < v_np, "Paramétrico", "No paramétrico")

  # Guardar en tablas
  resultados <- rbind(resultados, data.frame(
    n = n,
    ER = round(ER, 5),
    Metodo_mas_eficiente = metodo
  ))

  varianzas <- rbind(varianzas, data.frame(
    n = n,
    v_p = v_p,
    v_np = v_np
  ))
}

}

```

```

resultados %>%
  kbl(
    caption = "Eficiencia Relativa (ER) en una Gamma del Bootstrap Parametrico vs
    No Parametrico",
    col.names = c("Tamano muestral (n)", "Eficiencia Relativa (ER)",
                 "Metodo mas eficiente"),
    align = "ccc",
    booktabs = TRUE
  ) %>%
  kable_styling(
    bootstrap_options = c("striped", "hover", "condensed"),
    full_width = FALSE,
    position = "center"
  )

```



```

# --- Graficar variación de las varianzas ---
varianzas_largo <- varianzas %>%
  pivot_longer(cols = c(v_p, v_np), names_to = "Metodo", values_to = "Varianza")

ggplot(varianzas_largo, aes(x = n, y = Varianza, color = Metodo, group = Metodo)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(
    title = "Comparación de Varianzas Bootstrap \n Distribución Gamma",
    subtitle = "Métodos paramétrico vs no paramétrico",
    x = "Tamaño muestral (n)",
    y = "Varianza del estimador",
    color = "Método"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold"),
    legend.position = "top"
  )

```

7.2. Punto 2

```

set.seed(12345)
# Función para calcular N
calcular_N <- function() {
  suma <- 0
  n <- 0
}

```

```

while (suma <= 1) {
  n <- n + 1
  suma <- suma + runif(1) # Genera una variable aleatoria uniforme
}
return(n)
}

# Estimación de  $E[N]$  con 100 valores
estimacion_100 <- mean(replicate(100, calcular_N()))

# Estimación de  $E[N]$  con 1000 valores
estimacion_1000 <- mean(replicate(1000, calcular_N()))

# Estimación de  $E[N]$  con 10000 valores
estimacion_10000 <- mean(replicate(10000, calcular_N()))

```

7.3. Punto 3

```

set.seed(123)

# 1. Medias reales de los 5 periodos
medias <- c(131.37, 132.37, 134.47, 135.50, 136.27)
periodos <- paste0("Periodo_", 1:5)

# Tamaño de muestra por periodo
n <- 30

# 2. Estimación de desviación estándar
sd_est <- 4

# 3. Media común bajo  $H_0$ 
mu0 <- mean(medias)

# 4. Función de simulación
sim_orden <- function() {
  muestras <- replicate(5, mean(rnorm(n, mean = mu0, sd = sd_est)))
  all(diff(muestras) > 0)
}

# 5. Simulación Monte Carlo
B <- 10000
resultados <- replicate(B, sim_orden())

```

```

# 6. p-valor
p_valor <- mean(resultados)

# --- TABLA BONITA ---
tabla_resultados <- data.frame(
  Estadística = c("Media común bajo H0", "Desviación estándar usada",
                  "Tamaño de muestra por período", "p-valor Monte Carlo"),
  Valor = c(mu0, sd_est, n, p_valor)
)

kable(tabla_resultados, caption = "Resultados del procedimiento Monte Carlo",
      digits = 4, align = "c")

```



```

data <- data.frame(
  Periodo = 1:5,
  Altura = medias
)

```



```

ggplot(data, aes(x = Periodo, y = Altura, group = 1)) +
  geom_point(color = "orange", size = 3) +
  geom_line(color = "orange", linewidth = 1) +
  labs(
    title = "Altura media del cráneo por período dinástico",
    x = "Período",
    y = "Altura media (mm)"
  ) +
  theme_minimal(base_size = 14)

```

7.4. Punto 4

```

set.seed(123) # para reproducibilidad

## 1. Probabilidades bajo H0 (proporcionales a días del mes)
teo.prob <- c(31,28,31,30,31,30,31,31,30,31,30,31)
p_null <- teo.prob / sum(teo.prob)

## 2. Muestra piloto: totales por mes (hombres + mujeres)
tot <- c(721, 830, 812, 856, 716, 790, 751, 740, 803, 789, 801, 830)
N_piloto <- sum(tot)
p_alt <- tot / N_piloto # distribución "verdadera" estimada

```

```

## 3. Función para estimar la potencia para un n dado
sim_power <- function(n, B = 2000, alpha = 0.05) {
  rejections <- logical(B)

  for (b in 1:B) {
    # muestra simulada bajo H1
    x <- as.vector(rmultinom(1, size = n, prob = p_alt))

    # prueba chi-cuadrado contra H0
    # rescale.p=TRUE no es necesario aquí porque p_null ya suma 1,
    # pero no hace daño incluirlo.
    p_val <- suppressWarnings(chisq.test(x, p = p_null, rescale.p = TRUE)$p.value)

    rejections[b] <- (p_val < alpha)
  }

  mean(rejections) # estimación de la potencia
}

## 4. Barrido de tamaños muestrales
n_grid <- seq(100, 4000, by = 50) # puedes ajustar el rango
power_n <- sapply(n_grid, sim_power)

## 5. n óptimo: mínima n con potencia 0.80
n_opt <- min(n_grid[power_n >= 0.80])

plot(n_grid, power_n, type = "b", xlab = "Tamaño muestral n",
      ylab = "Potencia estimada",
      main = "Curva de potencia para prueba de uniformidad")
abline(h = 0.80, lty = 2, col = "red")
abline(v = n_opt, lty = 2, col = "blue")

```

7.5. Punto 5

7.5.1. Metodos

```

## =====
## 1. MÉTODOS DE GENERACIÓN
## =====

# Método 1: Aproximación por Newton-Raphson (del PDF)
normal_aprox <- function(u, error = 1e-6) {

```

```

x0 <- 0
err <- 1
while (err > error) {
  x1 <- x0 - (pnorm(x0) - u) / dnorm(x0)
  err <- abs(x1 - x0)
  x0 <- x1
}
return(x0)
}

# Método 2: Algoritmo de Greenberg (accept-reject)
normal_greenberg <- function() {
  repeat {
    u1 <- runif(1)
    u2 <- runif(1)
    u3 <- runif(1)

    x <- -log(u1) # exponencial

    # condición de aceptación
    if (u2 <= exp(-0.5*(x-1)^2)) {
      if (u3 <= 0.5) return(x)
      else return(-x)
    }
  }
}

```

7.5.2. Muestras

```

## =====
## 2. GENERACIÓN DE MUESTRAS
## =====

n <- 10000

time_NR <- system.time(x_NR <- sapply(runif(n), normal_aprox))[3] # Método 1
time_G <- system.time(x_G <- replicate(n, normal_greenberg()))[3] # Método 2
time_R <- system.time(x_R <- rnorm(n))[3] # Método 3

```

7.5.3. Pruebas

```
## =====
## 3. PRUEBAS DE NORMALIDAD
## =====

# KS test
KS_NR <- ks.test(x_NR, "pnorm")$p.value
KS_G <- ks.test(x_G, "pnorm")$p.value
KS_R <- ks.test(x_R, "pnorm")$p.value

# Shapiro-Wilk (muestras 5000)
SW_NR <- shapiro.test(sample(x_NR, 5000))$p.value
SW_G <- shapiro.test(sample(x_G, 5000))$p.value
SW_R <- shapiro.test(sample(x_R, 5000))$p.value

## =====
## 4. TABLA FINAL
## =====

tabla <- data.frame(
  Metodo = c("Newton-Raphson", "Greenberg", "rnorm"),
  Tiempo_seg = c(time_NR, time_G, time_R),
  KS_pvalue = c(KS_NR, KS_G, KS_R),
  SW_pvalue = c(SW_NR, SW_G, SW_R)
)

kable(tabla, digits = 5,
      caption = "Comparación de los tres generadores de la Normal Estándar")
```

7.5.4. QQPLOTS

```
## =====
## 4. GRAFICAS## =====

par(mfrow=c(1,3))
qqnorm(x_NR, main = "Normal aprox (NR)"); qqline(x_NR, col="red")
qqnorm(x_G, main = "Greenberg"); qqline(x_G, col="red")
qqnorm(x_R, main = "rnorm"); qqline(x_R, col="red")
```

Bioblografía

Luque-Calvo, P.L. (2017). *Escribir un Trabajo Fin de Estudios con R Markdown*. Disponible en <http://destio.us.es/calvo>.

Xie, Y., Dervieux, C. & Riederer, E. (2020). *R Markdown Cookbook*. Chapman; Hall/CRC, Boca Raton, Florida.