

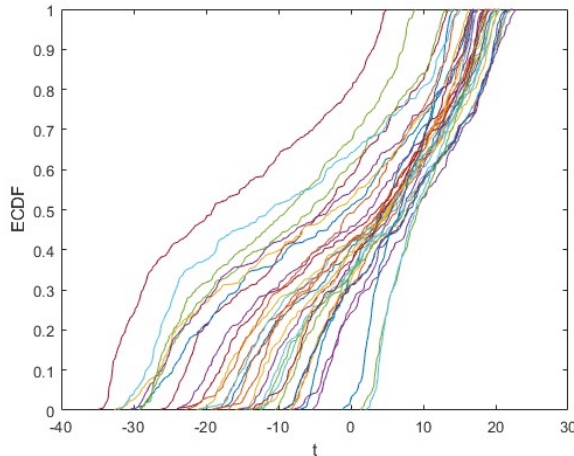
Taller 1 - Estadística No Paramétrica para Ciencia de los Datos

Abelino Sepúlveda Estrada
Maria Alejandra Vélez Clavijo
Alejandra Palacio Jaramillo
Ingeniería Matemática, Universidad EAFIT

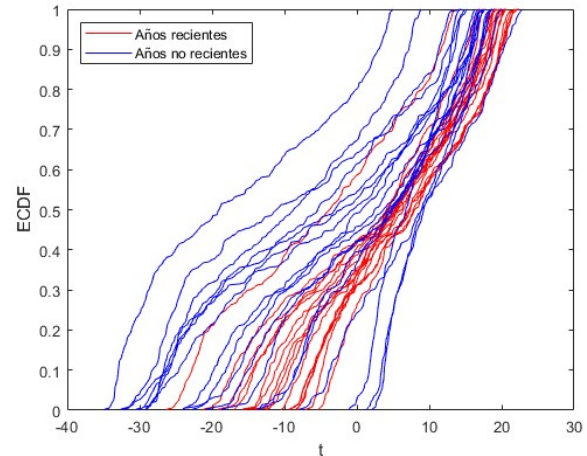
October 20, 2022

Punto 1

Grafique en un mismo plano las funciones de distribución empíricas de las temperaturas. Explique con base en las gráficas de las funciones, si es observable un efecto de cambio climático. Realice una tabla de doble entrada donde se muestre en cada par de años, cuál es más caluroso entre el par. Primero en términos de la temperatura media y otra tabla en términos del orden dado por la distribución empírica.



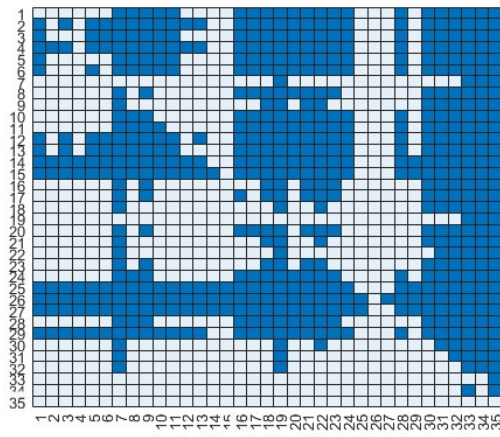
((a)) Distribución empírica temperaturas



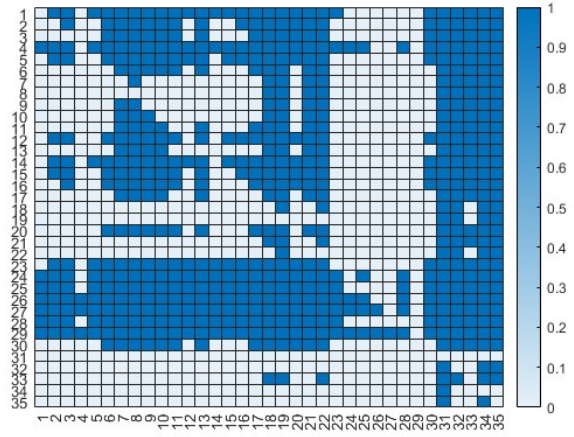
((b)) Análisis cambio climático

Figure 1: Función de Distribución Empírica

En la Figura 1, podemos evidenciar las funciones de distribución empírica de las temperaturas de cada año en Canadá. Adicional, en la Figura 1 (b). podemos observar la distribución empírica de las temperaturas en las ultimos 17 años (rojo) y los primeros 18 años (azul), donde es observable efectos de cambio climático, es decir, el nivel de temperatura en los años recientes es mayor respecto a los otros. Es decir, observamos efectos del calentamiento global.



((a)) Año más caliente según media



((b)) Año más caliente según la empírica

Figure 2: Función de Distribución Empírica

En la Figura 2 podemos ver la tabla de doble entrada donde se compara qué año es más caliente según la media y la distribución empírica. Donde el color azul indica que ese año es más caliente. También podemos ver que aunque ambas figuras sean muy parecidas, se evidencia que hay unas pequeñas diferencias entre algunos años.

Punto 2

Calcule el área limitada, en el primer cuadrante, por cada curva empírica y la recta $y = 1$. Réstele al área limitada, en el segundo cuadrante por cada curva y la recta $y = 0$. Esta área final sería un estimador Plug in de qué parámetro? Justifique.

Recordemos que la función de distribución empírica está acotada entre 0 y 1 en todo su dominio, entonces, esta diferencia es calculada con la siguiente integral:

$$\begin{aligned}
 \int_0^\infty [1 - F(t)] dt - \int_{-\infty}^0 F(t) dt &= \int_0^\infty [F(x)|_t^\infty] dt - \int_{-\infty}^0 [F(x)|_{-\infty}^t] dt \\
 &= \int_0^\infty \int_t^\infty f(x) dx dt - \int_{-\infty}^0 \int_{-\infty}^t f(x) dx dt \\
 &= \int_0^\infty \int_0^x f(x) dt dx + \int_{-\infty}^0 \int_x^0 -f(x) dt dx \\
 &= \int_0^\infty x f(x) dx + \int_{-\infty}^0 x f(x) dx \\
 &= \int_{-\infty}^\infty x f(x) dx
 \end{aligned}$$

Para determinar de qué estimador Plug-in sería este parámetro. Recordemos veamos qué es el Plug-in stimador

El estimador Plug-in de $\theta = T(F)$ está definido por

$$\hat{\theta}_n = T(\hat{F}_n) \quad (1)$$

Donde $T(F)$ es un funcional estadístico de una función F . Uno de los ejemplo es para la media, donde $\mu = \int x dF(x)$, así

$$T(\hat{F}_n) = \int a(x) d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n a(X_i) \quad (2)$$

Finalmente, graficamos el estimador Plug-in

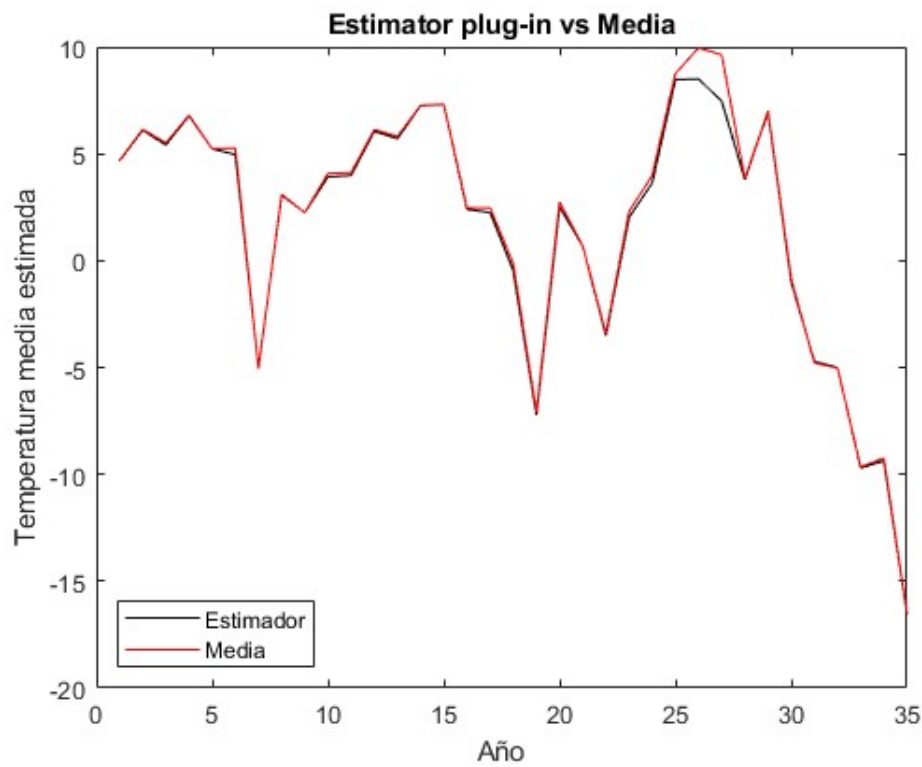


Figure 3: Caption

En la Figura 3 podemos evidenciar la comparación entre el Plug-in estimator y la media. Por lo que queda evidenciado que era el estimador Plug-in de la media.

Punto 3

Calcule y grafique las bandas de confianza a un 95% de confianza para la función de distribución empírica de los años en media más y menos calurosos. Hay sectores solapados? Argumente.

Sea una muestra de tamaño n y $1 - \alpha$ el nivel de confianza de las bandas. Los bandas de confianza no paramétrica se definen de la siguiente manera

$$L(x) = \max\{\hat{F}_n(x) - \epsilon_n, 0\} \quad (3)$$

$$U(x) = \min\{\hat{F}_n(x) + \epsilon_n, 1\} \quad (4)$$

donde

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}$$

Entonces, para todo F y para todo n

$$P(L(x) \leq F(x) \leq U(x), \quad \forall x) \geq 1 - \alpha \quad (5)$$

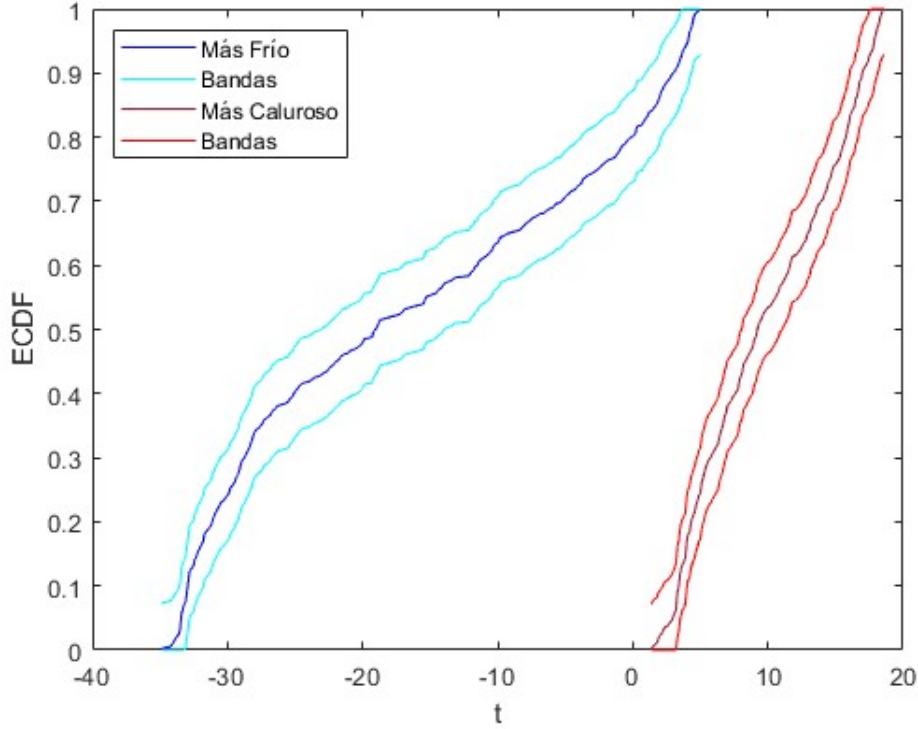


Figure 4: Bandas de confianza a un 95%

En la Figura 4 se puede observar que se cumple lo planteado en la Ecuación (5), es decir, que la función de distribución empírica de los años en media más y menos calurosos se encuentra completamente contenida en sus intervalos de confianza. Sin embargo, podemos ver que hay dos puntos donde se solapan en ambas funciones: en la banda inferior al principio y en la banda superior al final. Para la banda inferior, esto se puede explicar porque esta banda debe tener el mismo valor de la función empírica ya que no puede bajar de cero, como se muestra en la Ecuación (3). Análogamente se cumple con la banda superior de acuerdo con la Ecuación (4).

Punto 4

Escriba y ejecute un código que permita visualizar el Teorema de Glivenko Cantelli para una distribución Weibull con parámetro de escala, la nota que el grupo de trabajo espera obtener en este curso, y parámetro de forma, la edad media del grupo sobre el cuadrado de la nota esperada.

Teorema de Glivenko-Cantelli: sean $X_1, \dots, X_n \sim F$ y sea \hat{F}_n la empírica de CDF. Luego:

$$\sup_x \left| \hat{F}_n(x) - F(x) \right| \xrightarrow{\text{a.s.}} 0.$$

En otras palabras, el teorema de Glivenko-Cantelli establece que a medida que el tamaño de la muestra n aumenta, la distribución empírica converge de manera casi segura a la distribución teórica de la población. Por esto, cuando n es muy grande, la diferencia entre la distribución empírica y la distribución teórica tiende a ser cero.

Para visualizar este teorema para una distribución Weibull se consideró como parámetro de escala $a = 4.5$ y como parámetro de forma $b = 1.3827$. En las figuras 5, 6 y 7 se puede observar la distribución empírica para diferentes valores de n y la distribución teórica Weibull con los parámetros a y b determinados anteriormente. A partir de ellas se puede afirmar que, a medida que n aumenta, la distribución empírica converge a la distribución teórica, puesto que, por ejemplo, para $n = 5000$ es casi igual.

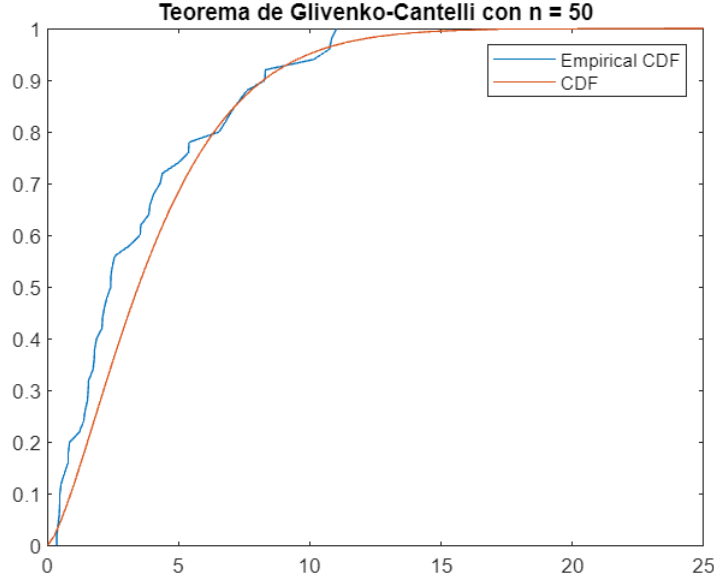


Figure 5: Aplicación del teorema de Glivenko Cantelli con $n = 50$

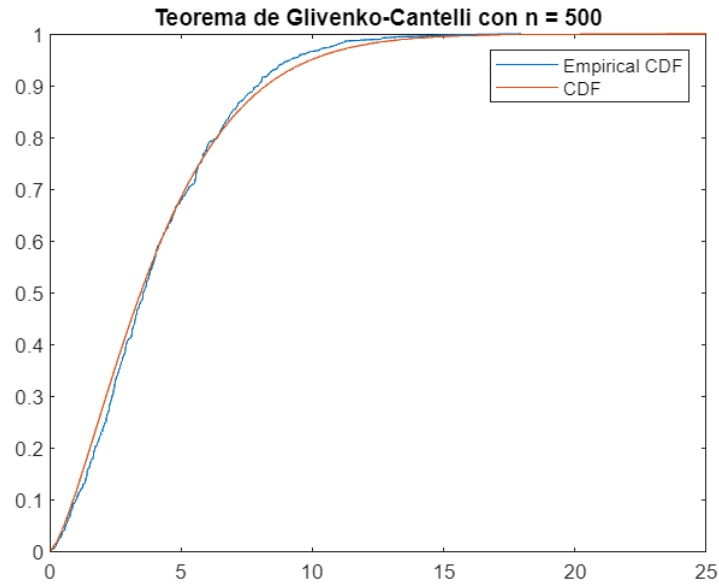


Figure 6: Aplicación del teorema de Glivenko Cantelli con $n = 500$

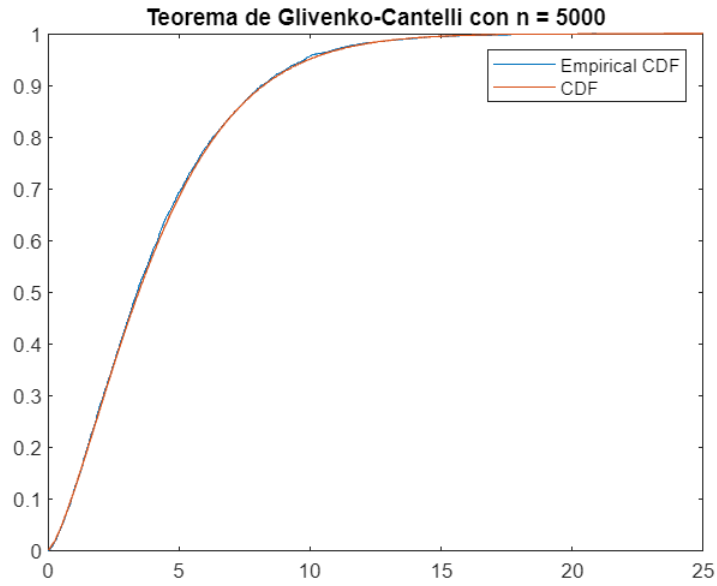


Figure 7: Aplicación del teorema de Glivenko Cantelli con $n = 5000$

Punto 5

Enuncie y demuestre la desigualdad de Jensen para funciones cóncavas. Muestre un ejemplo de utilidad de la desigualdad de Jensen para funciones cóncavas.

Sea X un variable aleatoria integrable y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ un función cóncava tal que $g(X)$ también es

integrable. Entonces se cumple la siguiente desigualdad, llamada la **Desigualdad de Jensen**

$$E[g(X)] \leq g(E[X])$$

Demostración: Una función g es convexa si, para cualquier punto, x_0 la gráfica de g se encuentra completamente por debajo de su tangente en el punto x_0 , es decir,

$$g(x) \leq g(x_0) + b(x - x_0), \quad \forall x$$

donde b es la pendiente de la tangente. Poniendo $x = X$ y $x_0 = E[X]$, la desigualdad se convierte en

$$g(X) \leq g(E[X]) + b(X - E[X])$$

Ahora, tomando el valor esperado de ambos lados de la desigualdad y tomando el hecho de que el valor esperado es un operador lineal, preserva las desigualdades, por lo que obtenemos

$$\begin{aligned} E[g(X)] &\leq E[g(E[X]) + b(X - E[X])] \\ &= E[g(E[X])] + bE[(X - E[X])] \\ &= E[g(E[X])] + b(E[X] - E[E[X]]) \end{aligned}$$

Por linealidad del valor esperado $E[E[X]] = E[X]$, así

$$\begin{aligned} E[g(X)] &\leq E[g(E[X])] + b(E[X] - E[X]) \\ &= E[g(E[X])] + b(0) \\ &= g(E[X]) \end{aligned}$$

Finalmente, una particular utilidad de la desigualdad de Jensen es la de los beneficios bancarios, como la de la de hipotecas de casas, ya que problemas como la volatilidad en el mercado de la vivienda puede llegar a causar problemas en los bancos. Matemáticamente (desigualdad de Jensen) esto puede ser explicado como: los beneficios bancarios son una función cóncava de los precios de vivienda. Las ganancias bancarias sería g , los precios de vivienda serían X . Así el promedio de $g(X)$, es decir $E[g(X)]$ es menor o igual que $g(\text{promedio de } X)$. En palabras, La ganancia promedio es menor o igual que la ganancia del promedio.

Punto 6

Consulte el concepto de L-statistics and M-statistics. Proporcione 4 ejemplos de cada uno. Exponga las ventajas de uno sobre el otro. Pinte la curva de su Influence function para cada uno de los ejemplos y obtenga conclusiones. Hable también del breakdown de cada uno de ellos.

L-estadísticos: En general, los L-estadísticos son aquellos que pueden ser escritos como una combinación lineal de estadísticos ordenados. En otras palabras, sea X_1, \dots, X_n una muestra de observaciones independientes que siguen una función de distribución F . La clase de L -estadísticos pueden ser presentados de la forma:

$$T_n = \sum_{i=1}^n c_{ni} h(X_{n:i}) + \sum_{j=1}^k a_j g(X_{n:[np_j]+1})$$

Donde c_{n1}, \dots, c_{nn} y a_1, \dots, a_k son coeficientes y h y g son funciones. Así, La primera está relacionada con los estadísticos ordenados y la segunda con la combinación lineal de algunos cuantiles. También, otra forma de representarlos es la siguiente:

$$\sum_{i=1}^n c_{n,i} F_n^{-1}(i/n)$$

Donde c_{n1}, \dots, c_{nn} son coeficientes y F_n^{-1} es la función de distribución inversa.

Algunos ejemplos de L-estadísticos son el primer estadístico ordenado (mínimo), el n-ésimo estadístico ordenado (máximo), la mediana y el rango, porque estos pueden expresarse como una combinación lineal de estadísticos ordenados.

El punto de ruptura (breakdown point) es el porcentaje máximo de valores extremos que soporta un estimador antes de no ser válido. Considerando los ejemplos anteriores, para el mínimo, el máximo y el rango el punto de ruptura es 0, lo cual significa que una sola observación puede cambiarlo arbitrariamente. Por otro lado, el punto de ruptura de la mediana es 0,5, lo cual confirma que es un estadístico robusto.

M-estadísticos: Los M-estadísticos están definidos por la solución de:

$$\hat{\theta} = \arg \min \theta \left(\sum i = 1^n \rho(x_i, \theta) \right)$$

el cual es un problema de minimización respecto a θ para cierta función ρ . Cabe destacar que esta definición de M-estadístico proviene de la generalización de la estimación de máxima verosimilitud.

En general, dentro de los ejemplos de los M-estadísticos se encuentran los estimadores de máxima verosimilitud y los mínimos cuadrados no lineales. Además, otros ejemplos asociados a esta clase de estadísticos podrían ser los siguientes dos casos, donde el problema se minimiza cuando θ es la media (breakdown point = 0) y la mediana (breakdown point = 0,5), respectivamente:

$$\hat{\theta} = \arg \min \theta \left(\sum i = 1^n \frac{(x - \theta)^2}{2} \right)$$

$$\hat{\theta} = \arg \min \theta \left(\sum i = 1^n |x - \theta| \right)$$

Es importante resaltar que la principal diferencia entre los L-estadísticos y los M-estadísticos es que los primeros permiten tener un acercamiento inicial a los datos, puesto que son fáciles de calcular. Sin embargo, no son robustos, a diferencia de los M-estadísticos, los cuales pueden brindar información más precisa acerca del conjunto de datos.

Referencias utilizadas para el desarrollo de este punto: [5], [1], [4], [2] y [3]

Punto 7

Deduzca la distribución y densidad del iésimo estadístico ordenado. Exponga con detalle cómo sería un procedimiento sencillo de simular el jésimo estadístico ordenado. Simule 1000 observaciones de

algún estadístico ordenado de una muestra de tamaño n que venga de una distribución Weibull. Pinte en un mismo plano su distribución empírica y su distribución teórica.

Deducción de la distribución y densidad del i -ésimo estadístico ordenado:

En primer lugar, deduzcamos la función de distribución. Sea $X_1, \dots, X_n \sim F$ una muestra de observaciones que siguen una función de distribución F . Ahora, es conveniente definir las variables aleatorias Z_i de la siguiente manera:

$$Z_i = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases}$$

Respecto a las Z_i podemos afirmar que son independientes e idénticamente distribuidas, las cuales siguen una distribución Bernoulli con parámetro $\theta = P(X_i \leq x) = F(x)$, es decir, $Z_1, \dots, Z_n \sim Ber(\theta)$. Ahora bien, es claro que $Z_1 + \dots + Z_n \sim Bin(n, \theta)$ y representan el número de variables aleatorias Z_i tales que $X_i \leq x$. A partir de esto, podemos plantear que la función de distribución del i -ésimo estadístico de orden está dada por:

$$F_{X_{(i)}}(x) = P(X_{(i)} \leq x)$$

La anterior expresión significa que cada Z_i toma un valor menor o igual a x , lo que en otras palabras significa que el número de variables aleatorias de la muestra que toman un valor menor o igual a x es mínimo i . Así, como la suma de las Z_i sigue una distribución binomial, tenemos:

$$F_{X_{(i)}}(x) = P(Y_1 + \dots + Y_n \geq i)$$

$$F_{X_{(i)}}(x) = \sum_{j=i}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}$$

Teniendo la función de distribución, para obtener la función de densidad del i -ésimo estadístico ordenado es necesario derivar a $F_{X_{(i)}}(x)$:

$$f_{X_{(i)}}(x) = \sum_{j=i}^n \binom{n}{j} j [F(x)]^{j-1} f(x) [1 - F(x)]^{n-j} - \sum_{j=i}^n \binom{n}{j} [F(x)]^j (n-j) [1 - F(x)]^{n-j-1} f(x)$$

$$f_{X_{(i)}} = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i} + \sum_{j=i+1}^n \binom{n}{j} j [F(x)]^{j-1} [1 - F(x)]^{n-j} f(x)$$

$$- \sum_{k=j}^{n-1} \binom{n}{j} (n-j) [F(x)]^j [1 - F(x)]^{n-j-1} f(x)$$

$$f_{X_{(i)}} = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i} + \sum_{j=i}^{n-1} \binom{n}{j+1} (j+1) [F(x)]^j [1 - F(x)]^{n-j-1} f(x)$$

$$-\sum_{k=j}^{n-1} \binom{n}{j} (n-j)[F(x)]^j [1-F(x)]^{n-j-1} f(x)$$

Ahora, la diferencia entre el segundo y el tercer término de la expresión es igual a cero, puesto que:

$$\begin{aligned} \frac{n!}{j!(n-j)!} (n-j) &= \frac{n!}{(j+1)!(n-j-1)!} (j+1) \\ \frac{n!}{j!(n-j)(n-j-1)!} (n-j) &= \frac{n!}{(j+1)j!(n-j-1)!} (j+1) \\ \frac{n!}{j!(n-j-1)!} &= \frac{n!}{j!(n-j-1)!} \end{aligned}$$

Finalmente, se puede concluir que la función de densidad del i -ésimo estadístico ordenado es igual a:

$$f_{X_{(i)}} = \binom{n}{i} i f(x) [F(x)]^{i-1} [1-F(x)]^{n-i}$$

Ahora bien, para simular el i -ésimo estadístico ordenado, en primer lugar se determina el tamaño n de la muestra (para el ejemplo planteado para este ejercicio, se escogió $n = 2000$). Luego, se simula una muestra aleatoria de n observaciones que siguen una distribución dada. Luego, se ordena el conjunto de datos para seleccionar el i -ésimo estadístico ordenado y se repite este proceso las veces deseadas. Finalmente, se calcula la distribución teórica y, para fines del ejercicio, se calcula la distribución empírica. Como ejemplo, se simularon 1000 muestras aleatorias que siguen una distribución Weibull con parámetro de escala 4.5 y parámetro de forma 1.3827, donde se seleccionó en cada una el i -ésimo estadístico ordenado (máximo) y se calculó la distribución empírica y la distribución teórica, la cual es $[F(t)]^n$, donde $F(t)$ es la distribución teórica de la Weibull. Las gráficas obtenidas son las que se muestran en la figura 8.

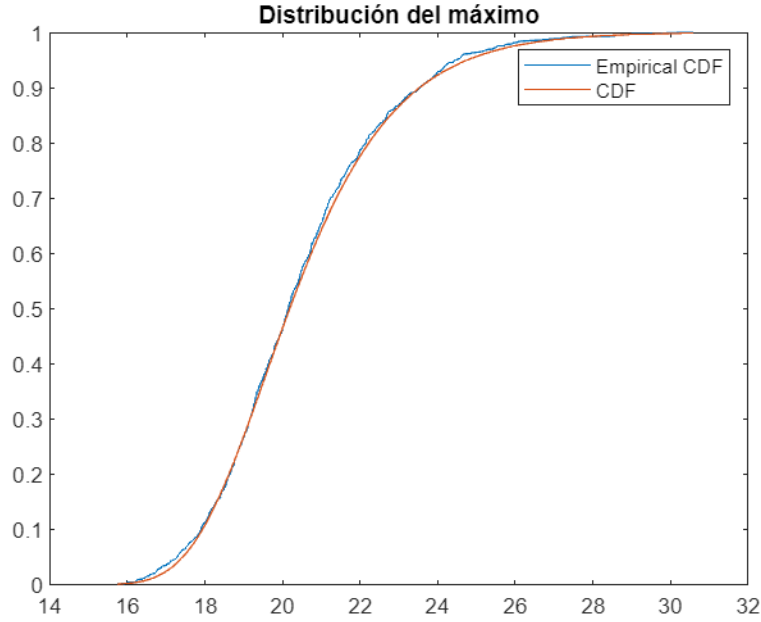


Figure 8: Distribución del máximo con $n = 2000$

Punto 8

Suponga que X es una variable aleatoria exponencial con parámetro β . Calcule

$$P(|X - \mu| > k\sigma)$$

Para $k > 1$. Compare el resultado con la cota obtenida de la desigualdad de Chebyshev. Recordemos la desigualdad de Chebyshev, sea X una variable aleatoria con $E[X] = \mu$ y $Var[X] = \sigma^2$, Para $k > 0$

$$\begin{aligned} P(|X - \mu| \geq k\sigma) &\leq \frac{\sigma^2}{(k\sigma)^2} \\ &= \frac{1}{k^2} \end{aligned}$$

Ahora, sea

$$\begin{aligned} P(|X - \mu| > k\sigma) &= P(-k\sigma > X - \mu > k\sigma) \\ &= P(-k\sigma + \mu > X > k\sigma + \mu) \\ &= P(X < \mu - k\sigma) + P(X > k\sigma + \mu) \\ &= P(X < \mu - k\sigma) + 1 - P(X < k\sigma + \mu) \\ &= P\left(X < \frac{1-k}{\beta}\right) + 1 - P\left(X < \frac{1+k}{\beta}\right) \end{aligned}$$

Como $k < 1$, así $1 - k < 0$, por lo que.

$$\begin{aligned}
P(|X - \mu| > k\sigma) &= 1 - P\left(X < \frac{1+k}{\beta}\right) \\
&= 1 - \left(1 - e^{-\frac{k+1}{\beta^2}}\right) \\
&= e^{-\frac{k+1}{\beta^2}} \\
&= P(|X - \mu| > k\sigma) \\
&\leq \frac{\sigma^2}{(k\sigma)^2} \\
&= \frac{1}{k^2}
\end{aligned}$$

Finalmente, podemos observar que $e^{-\frac{k+1}{\beta^2}}$ es siempre menor que 1 porque $k > 1$ y por este mismo motivo, $\frac{1}{k^2}$ también es menor que 1. Por lo tanto,

$$P(|X - \mu| > k\sigma)$$

es acotado por la desigualdad de Chebyshev.

Punto 9

Demuestre que si X es Poisson de parámetro λ entonces

$$P(X \geq 2\lambda) \leq \frac{1}{\lambda}$$

Sea $X \sim \text{Poisson}(\lambda)$. Recordando la desigualdad de Chebyshev, con $\epsilon > 0$ y X una variable aleatoria con $E[X] = \mu$ y $\text{Var}[X] = \sigma^2$, entonces

$$P[|X - \mu| \geq \epsilon] \leq \frac{\sigma^2}{\epsilon^2}$$

Como $X \sim \text{Poisson}(\lambda)$, $E[X] = \text{Var}[X] = \lambda$, si establecemos $\epsilon = \lambda$:

$$\begin{aligned}
P(|X_\lambda| \geq \lambda) &\leq \frac{1}{\lambda} \\
P(-\lambda \geq X - \lambda \geq \lambda) &\leq \frac{1}{\lambda} \\
P(-\lambda + \lambda \geq X \geq \lambda + \lambda) &\leq \frac{1}{\lambda} \\
P(0 \geq X \geq 2\lambda) &\leq \frac{1}{\lambda} \\
P(X \leq 0) + P(X \geq 2\lambda) &\leq \frac{1}{\lambda} \\
P(X \geq 2\lambda) &\leq \frac{1}{\lambda}
\end{aligned}$$

Punto 10

Demuestre convergencia en probabilidad está implicada por la convergencia en media cuadrática.

Definamos convergencia en probabilidad y en media cuadrática.

Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias con distribución $F_n(x)$. Sea X una variable aleatoria con distribución $F(x)$

$$X_n \xrightarrow{p} X \iff \forall \alpha > 0 \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \alpha) = 0$$

Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias con distribución $F_n(x)$. Sea X una variable aleatoria con distribución $F(x)$

$$X_n \xrightarrow{q.m.} X \iff \lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$$

Ahora, provemos que $X_n \xrightarrow{q.m.} X \implies X_n \xrightarrow{p} X$.

$$\begin{aligned} &= P[|X_n - X|^2 > \alpha^2] \\ 0 \leq P[|X_n - X| > \alpha] &\leq \frac{E(X_n - X)^2}{\alpha^2} \quad \text{desigualdad de markov} \end{aligned}$$

$$0 \leq \lim_{n \rightarrow \infty} P[|X_n - X| > \alpha] \leq \lim_{n \rightarrow \infty} \frac{E(X_n - X)^2}{\alpha^2}$$

Y como converge en media cuadrática $\lim_{n \rightarrow \infty} \frac{E(X_n - X)^2}{\alpha^2} = 0$ por lo que,

$$\begin{aligned} 0 \leq \lim_{n \rightarrow \infty} P[|X_n - X| > \alpha] &\leq 0 \\ \lim_{n \rightarrow \infty} P[|X_n - X| > \alpha] &= 0 \end{aligned}$$

Por lo tanto, $X_n \xrightarrow{q.m.} X \implies X_n \xrightarrow{p} X$.

Punto 11

Demuestre que la función de distribución converge en probabilidad a la distribución teórica.

Sea $\{X_n\}_{n \in \mathbb{N}}$ variables aleatorias reales independientes e idénticamente distribuidas, la distribución de distribución empírica \hat{F}_n se define

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$$

donde

$$I(X_i \leq x) = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{en otro caso} \end{cases}$$

Ahora, provemos que $\hat{F}_n \xrightarrow{p} F$. Para esto veamos que el $\lim_{n \rightarrow \infty} \text{MSE}(\hat{F}_n(x)) = 0$, es decir, para un n demasiado grande, el error cuadrático medio de la función de distribución empírica es cero (definición de convergencia en probabilidad).

Veamos la definición de MSE :

$$MSE(\hat{F}_n(x)) = Var[\hat{F}_n(x)] + Bias(\hat{F}_n(x), F(x))$$

Ahora, calculemos el $Bias$ /sesgo.

$$\begin{aligned} E[\hat{F}_n(x)] &= E\left[\frac{1}{n} \sum_{i=1}^n I(X_i \leq x)\right] \\ &= \frac{1}{n} \sum_{i=1}^n E[I(X_i \leq x)] \\ &= \frac{1}{n} \sum_{i=1}^n P(X_i \leq x) \\ &= \frac{1}{n} \sum_{i=1}^n F(x) \\ &= F(x) \end{aligned}$$

Por lo que $Bias(\hat{F}_n(x), F(x)) = 0$

Procedemos a calcular el error cuadrático medio

$$\begin{aligned} MSE(\hat{F}_n(x)) &= Var[\hat{F}_n(x)] + Bias(\hat{F}_n(x), F(x)) \\ &= Var[\hat{F}_n(x)] \\ &= Var\left[\frac{1}{n} \sum_{i=1}^n I(X_i \leq x)\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n Var[I(X_i \leq x)] \\ &= \frac{1}{n^2} \sum_{i=1}^n P(X_i \leq x)(1 - P(X_i \leq x)) \\ &= \frac{1}{n^2} \sum_{i=1}^n F(x)(1 - F(x)) \\ &= \frac{1}{n} F(x)(1 - F(x)) \end{aligned}$$

Así,

$$\begin{aligned} \lim_{n \rightarrow \infty} MSE(\hat{F}_n(x)) &= \lim_{n \rightarrow \infty} \frac{1}{n} F(x)(1 - F(x)) \\ &= 0 \end{aligned}$$

Por lo que finalmente queda demostrado que $\hat{F}_n \xrightarrow{p} F$

Punto 12

Considere las temperaturas diarias del año en media más frío. Calcule un intervalo de confianza con la temperatura para la temperatura máxima. Calcule el sesgo de $T_{[n]}$ y la varianza.

Para calcular el intervalo de confianza de $T_{[n]}$ lo hacemos por medio de Bootstrap de la siguiente forma

$$[T_n - z_{\alpha/2} \hat{se}_{boot} \quad T_n + z_{\alpha/2} \hat{se}_{boot}]$$

donde $\hat{se}_{boot} = \sqrt{v_{boot}}$ es la estimación por bootstrap del error estándar y para calcular v_{boot} hacemos B remuestreos con repetición y calculamos el estadístico deaseado, así obtenemos

$T_{n,1}^*, T_{n,2}^*, \dots, T_{n,B}^*$ y sea

$$v_{boot} = \frac{1}{B} \sum_{b=1}^B \left(T_{n,b}^* - \frac{1}{B} \sum_{r=1}^B T_{n,r}^* \right)^2 \quad (6)$$

Finalmente, el intervalo de confianza para $T_{[n]} = 5$ para un $\alpha = 0.05$ y un $B = 1000$

$$[4.5 \quad 5]$$

Ahora, para calcular el sesgo se propone usar jackknife, en el cual se puede calcular el sesgo del parámetro sin conocer el estimador de la siguiente manera

$$b_{jack} = (n-1) * (\bar{T}_n - T_n)$$

donde $\bar{T}_n = \frac{1}{n} \sum_{i=1}^n T_{(-i)}$ y $T_{(-i)}$ es el estadístico con la i -ésima observación removida. Así

$$b_{jack} = -0.0997$$

Finalmente, hacemos la estimación de la varianza por medio de la ecuación (6).

$$v_{boot} = 0.0140$$

Punto 13

Considere U_1, U_2, \dots, U_n una muestra de una distribución uniforme en el intervalo $[0, 1]$. Calcule la distribución teórica de $U_{[1]}$, su media y sesgo. Genere la muestra y utilice bootstrap para calcular la varianza de $U_{[1]}$. Calcule el sesgo por Jackknife y compárelo con el sesgo teórico.

En primer lugar, recordemos que la función de distribución del i -ésimo estadístico ordenado está dada por:

$$f_{X(i)} = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}$$

$$f_{X(i)} = \frac{n!}{i!(n-i)!} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}$$

$$f_{X(i)} = \frac{n!}{(i-1)!(n-i)!} f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}$$

En este caso, para la muestra dada y, teniendo en cuenta que la función de distribución para una uniforme está dada por $F(x) = \frac{x-a}{b-a}$ y la función de densidad por $f(x) = \frac{1}{b-a}$, se consigue que la función de distribución para el primer estadístico ordenado es:

$$f_{X_{(1)}} = \binom{n}{1} (1) \frac{1}{b-a} \left[\frac{x-a}{b-a} \right]^{1-1} \left[1 - \frac{x-a}{b-a} \right]^{n-1}$$

$$f_{X_{(1)}} = \binom{n}{1} [1-x]^{n-1}$$

$$f_{X_{(1)}} = \frac{n!}{(1-1)!(n-1)!} [1-x]^{n-1}$$

$$f_{X_{(1)}} = n [1-x]^{n-1}$$

Ahora, analicemos la forma de la expresión de la cual se obtiene la función de distribución. Dado que la muestra aleatoria proviene de una distribución $U(0, 1)$, entonces la distribución de probabilidad del i -ésimo estadístico ordenado es una distribución $B(i, n-i+1)$. Así, la media del i -ésimo estadístico ordenado y particularmente del primer estadístico ordenado es:

$$E[X_{(i)}] = \frac{i}{n+1}$$

$$E[X_{(i)}] = \frac{1}{n+1}$$

Con esto determinado, el sesgo del primer estadístico ordenado es insesgado, pues:

$$bias(U_{[1]}) = \frac{1}{n+1}$$

Para el desarrollo del ejercicio propuesto, generamos una muestra aleatoria con 5000 observaciones y utilizamos bootstrap para calcular la varianza de $U_{[1]}$, la cual dió como resultado $5.1423e-08$. Luego, se calculó el sesgo teórico (aplicando la fórmula dada anteriormente), el cual fue $1.9996e-04$ y el sesgo por Jackknife, el cual fue de $3.7219e-04$. De estos resultados, se puede concluir que ambos son prácticamente cero, por tanto, demuestran que el estimador es insesgado.

Punto 14

Exponga las diferencias entre el bootstrap paramétrico y no paramétrico, ilustre con ejemplos. Consulte versiones robustas del método bootstrap y exponga ejemplos sobre el performance.

Bootstrap paramétrico y no paramétrico: En general, Bootstrap es un método estadístico que usualmente se usa para la estimación de la varianza y la distribución de un estadístico. Además, con este método también se pueden construir intervalos de confianza. Respecto al bootstrap paramétrico y no paramétrico existe una diferencia clave, la cual radica en la manera en la cual se calculan las distribuciones de las cuales vamos a sacar las muestras para el Bootstrap. Por ejemplo, si F_θ depende del parámetro θ y $\hat{\theta}$ es el estimador de θ , entonces muestreamos a partir de $F_{\hat{\theta}}$ y, como estamos dependiendo de la estimación de un parámetro, nos referimos a la

versión paramétrica del Bootstrap. Por otro lado, cuando hablamos de bootstrap no paramétrico, significa que las muestras serán obtenidas a partir de la distribución empírica \hat{F}_n , que viene a ser el estimador no paramétrico de F (?).

En el ejemplo mostrado en la figura 9 se muestran histogramas correspondientes al bootstrap paramétrico y no paramétrico. El tamaño de la muestra simulada es de 200 y son datos provenientes de una normal con media 10 y desviación estándar 2.

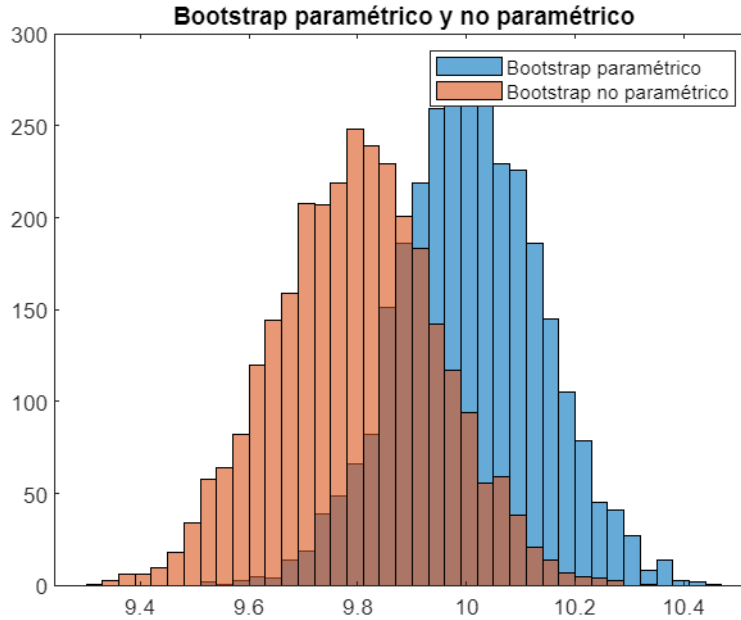


Figure 9: Bootstrap con $n = 200$

Punto 15

Explique una forma no paramétrica y robusta de calcular la distancia de Mahalanobis. Aplique la técnica para recortar la muestra bivalente de temperaturas correspondientes a los años, en media, más caluroso y más frío. Compáre con el recorte obtenido de forma habitual. Luego elija aleatoriamente 30 observaciones y súmele un ruido que venga de una normal con media 10y varianza 0.25y realice de nuevo el ejercicio.

La distancia de Mahalanobis entre dos variables aleatorias con la misma distribución de probabilidad \vec{x} y \vec{y} con matriz de covarianza Σ se define como:

$$d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})}$$

Una forma no paramétrica y robusta de calcular la distancia de Mahalanobis es interviniendo la matriz de covarianza, específicamente el coeficiente de correlación, puesto que este se calcula por máxima verosimilitud, lo cual lo hace sensible a datos sensibles respecto a otros coeficientes de correlación. En este caso, se utilizará el coeficiente de correlación de Kendall para llevar la distancia de Mahalanobis tradicional a una versión robusta. Dado que la covarianza entre x y y puede definirse como:

$$\text{Cov}(x, y) = \rho_k S_x S_y$$

Donde ρ_k es el coeficiente de correlación de Kendall, S_x es la desviación estándar de la característica x y S_y es la desviación estándar de la característica y . Asimismo, el coeficiente de correlación de Kendall se define como:

$$p_k = \frac{(\# \text{ de pares concordantes}) - (\# \text{ de pares discordantes})}{\binom{n}{2}}$$

Cuando cualquier par de observaciones (x_i, y_i) y (x_j, y_j) , donde $i < j$ se dice que son un par concordante si ambos $x_i > x_j$ y $y_i > y_j$ o el caso de que $x_i < x_j$ y $y_i < y_j$, de no ocurrir esto se dice que el par es discordante.

Ejercicio 2 - Página 10

(Experimento computacional) Compare la cobertura y la longitud de

$$\hat{p}_n \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \quad (1.18)$$

y

$$\hat{p}_n \pm \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)} \quad (1.19)$$

por simulación. Donde

$$\hat{p}_n = n^{-1} \sum_{i=1}^n X_i$$

con $X_i \sim \text{Bernoulli}(p)$. Tome $p = 0,2$ y utilice $\alpha = 0.05$. Pruebe con varios tamaños de muestra n . ¿Qué tamaño debe tener n para que el intervalo puntual tenga una cobertura precisa? ¿Cómo se comparan las longitudes de los dos intervalos cuando se alcanza este tamaño de la muestra?

En el experimento computacional realizado se tomaron valores de n desde 10 hasta 100.000 con un tamaño de paso de 100. Para conocer la cobertura de los intervalos e identificar el tamaño de n para que los intervalos muestren una cobertura precisa primero se realizaron gráficas para analizar el comportamiento de los intervalos a medida de que n crece. Tal como se muestra en Figura 10, el intervalo (1.18) tiene mejor comportamiento, puesto que a medida que n crece, este se hace cada vez más pequeño centrando su valor en p , y lo hace con mayor rapidez que (1.19).

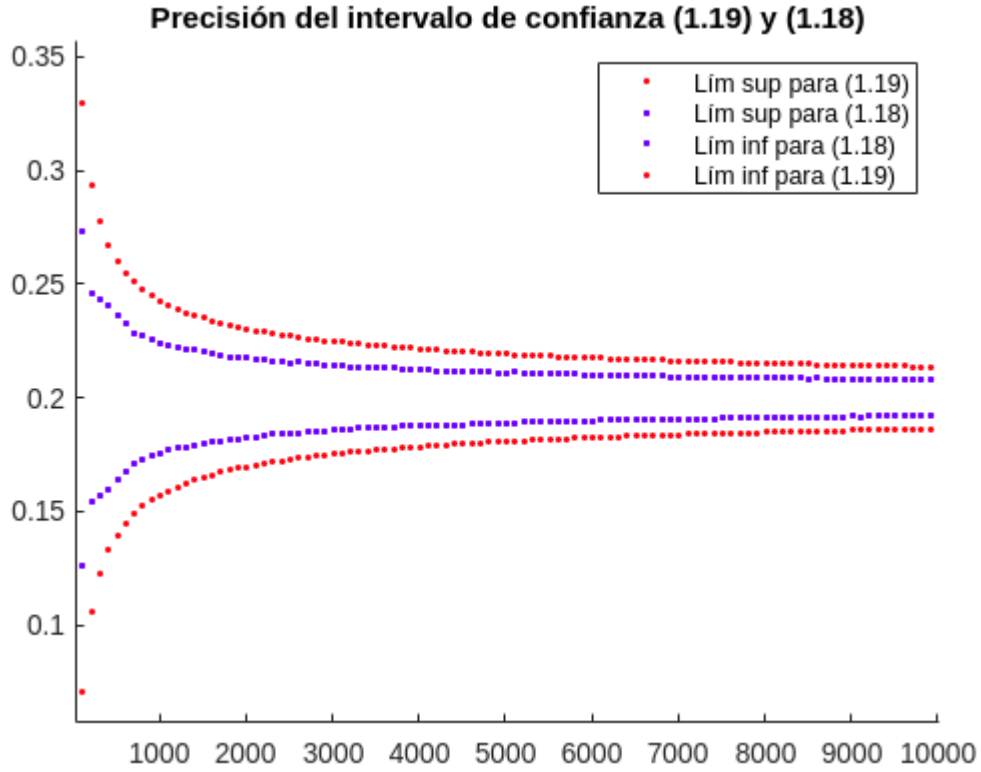


Figure 10: Cobertura de los intervalos de confianza (1.18) y (1.19)

También se observó el comportamiento en cuanto a la diferencia en el límite del intervalo a medida que n crece, con el objetivo de identificar si existe convergencia en la cobertura cuando se tiene a n grande. En efecto, tanto en (1.18) como en (1.19) tal como vemos en Figura 11, a partir de cierto n , la diferencia entre estos límites cada vez se hace más cercana a cero. Dado que las funciones que definen a estos intervalos no son completamente decrecientes, se observan leves fluctuaciones. Sin embargo, fue posible identificar que para $n = 100$ teníamos diferencias de orden 10^{-2} , con $n = 210$, 10^{-3} , $n = 1300$, 10^{-4} , y para $n = 3400$, 10^{-5} . Es decir, a partir de $n = 200$ ya ambos intervalos tienen cobertura bastante buena.

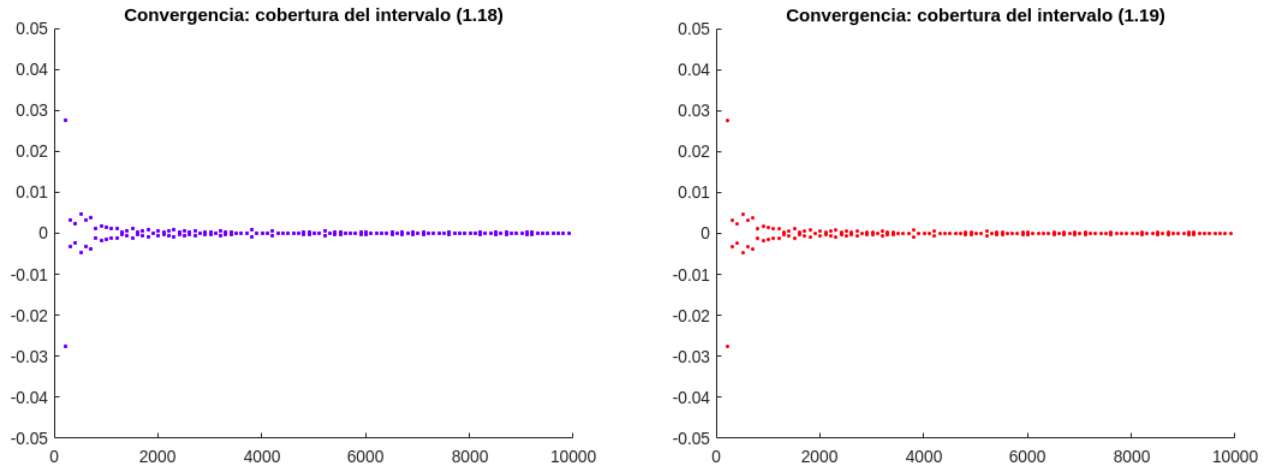


Figure 11: Convergencia: cobertura de los intervalos.

En la Figura 12 se realiza una comparación de la longitud de los intervalos a partir de $n = 210$. La gráfica resultante permite concluir que desde $n = 210$ en adelante, el intervalo (1.18) tiende a tener una longitud menor que la del intervalo (1.19), por lo tanto, su cobertura es más precisa. Sin embargo, el comportamiento en 1.19 es más constante que en 1.18, ya que en este último se muestran leves variaciones en el comportamiento con el que decrece.

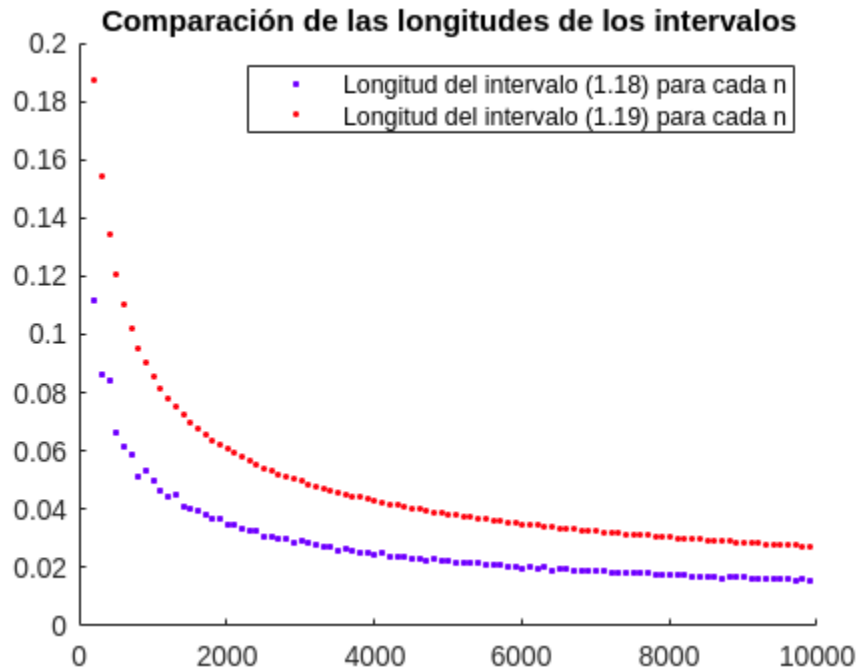


Figure 12: Comparación de las longitudes de los intervalos

Ejercicio 3 - Página 24

(Experimento computacional) Genere 100 observaciones de una distribución $N(0, 1)$. Compute las bandas de confianza a un 95% para la CDF F . Repítalo 1000 veces y mire qué tan seguido las bandas contienen la verdadera función de distribución. Repítalo usando datos de una distribución Cauchy.

Para hacer la simulación, se generaron 100 datos de una distribución uniforme 1000 veces y posteriormente se generaron las bandas de confianza. Finalmente, se obtuvo que el 98.2% de las veces las bandas contienen la verdadera función de distribución.

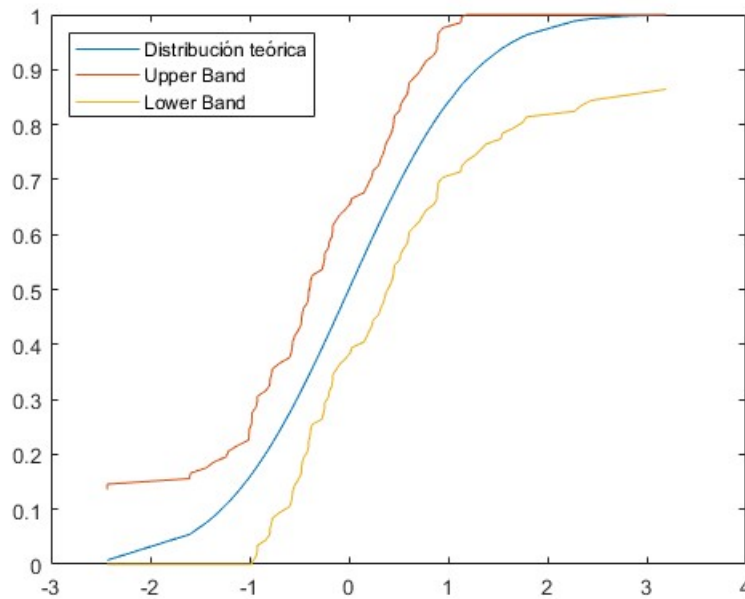


Figure 13: Última simulación bandas de confianza Normal estándar

En la Figura 13 podemos ver los resultados de la última simulación de las observaciones normales, donde se evidencia que la distribución teórica está completamente contenida en las bandas de confianza.

Análogamente, se realizó el proceso para la distribución cauchy. Se obtuvo que en un 95.2% de las simulaciones las bandas contenían la verdadera función de distribución cauchy.

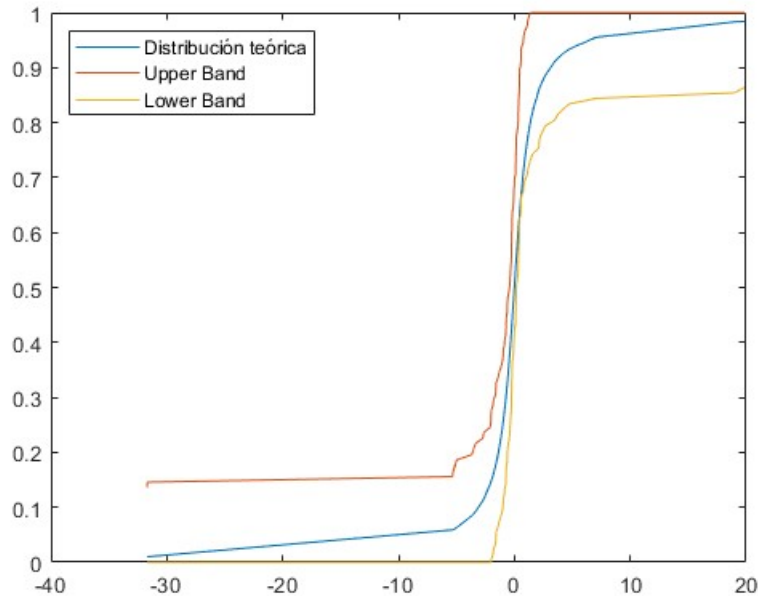


Figure 14: Última simulación bandas de confianza Cauchy

Ejercicio 2 - Página 39

LSAT	576	635	558	578	666	580	555	661
	651	605	653	575	545	572	594	
GPA	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43
	3.36	3.13	3.12	2.74	2.76	2.88	3.96	

Figure 15: Puntuaciones LSAT y GPA

Cada punto de datos es de la forma $X_i = (Y_i, Z_i)$ donde $Y_i = LSAT_i$ y $Z_i = GPA_i$. Encuentre la estimación plug-in del coeficiente de correlación. Estime el error estándar utilizando (i) la función de influencia, (ii) el jackknife y (iii) el bootstrap. A continuación, calcule un intervalo de confianza bootstrap pivotado al 95 por ciento. Tendrá que calcular el error estándar de T^* para cada muestra bootstrap.

0.5459

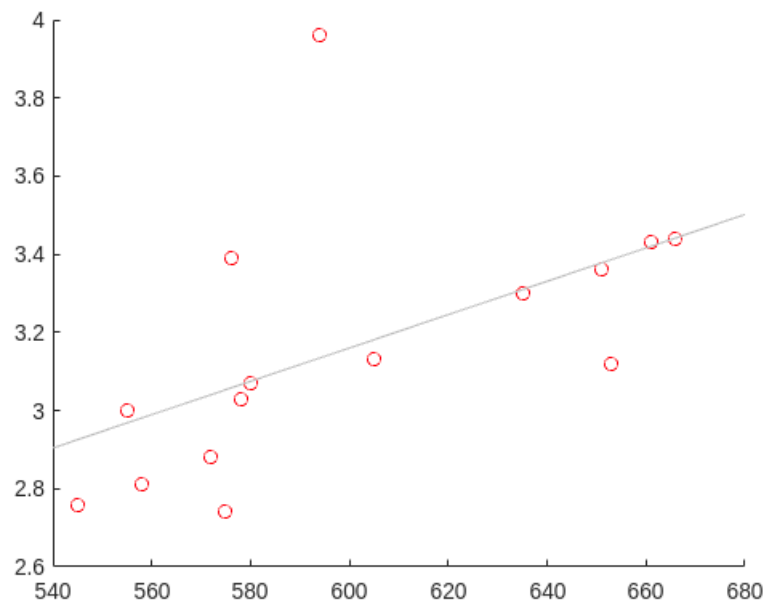


Figure 16: Línea de mínimos cuadrados entre LSAT y GPA

La estimación plug-in del coeficiente de correlación obtuvo como resultado **0.5459**. Además, como se puede observar en la Figura 20, existe una correlación positiva entre LSAT Y GPA. Realizamos bootstrap para calcular la significancia estadística.

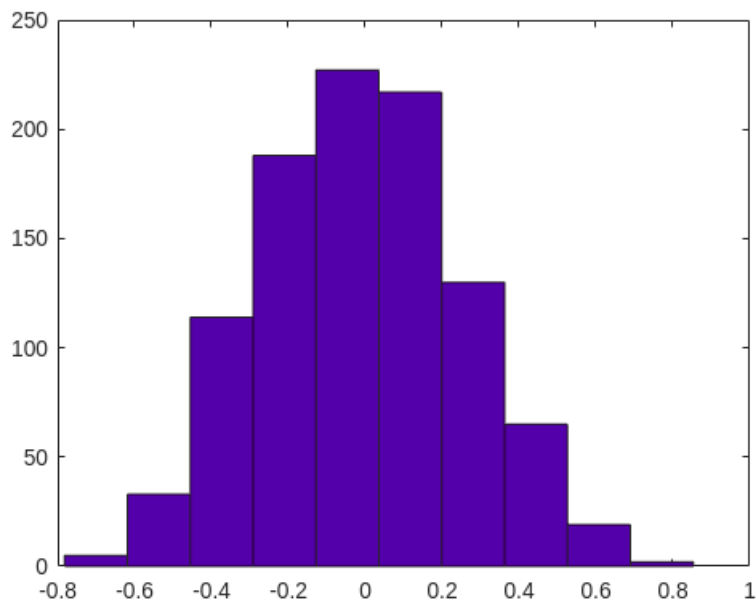


Figure 17: Histograma del bootstrap, remuestreo de 1000 para la correlación entre LSAT y GPA

Los siguientes datos fueron utilizados para ilustrar el bootstrap por Bradley Efron, el inventor del bootstrap. Los datos son las puntuaciones del LSAT (para para entrar

en la facultad de derecho) y el GPA.

LSAT	576	635	558	578	666	580	555	661
	651	605	653	575	545	572	594	
GPA	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43
	3.36	3.13	3.12	2.74	2.76	2.88	3.96	

Figure 18: Puntuaciones LSAT y GPA

Cada punto de datos es de la forma $X_i = (Y_i, Z_i)$ donde $Y_i = LSAT_i$ y $Z_i = GPA_i$. Encuentre la estimación plug-in del coeficiente de correlación. Estime el error estándar utilizando (i) la función de influencia, (ii) el jackknife y (iii) el bootstrap. A continuación, calcule un intervalo de confianza bootstrap pivotado al 95 por ciento. Tendrá que calcular el error estándar de T^* para cada muestra bootstrap. 0.5459

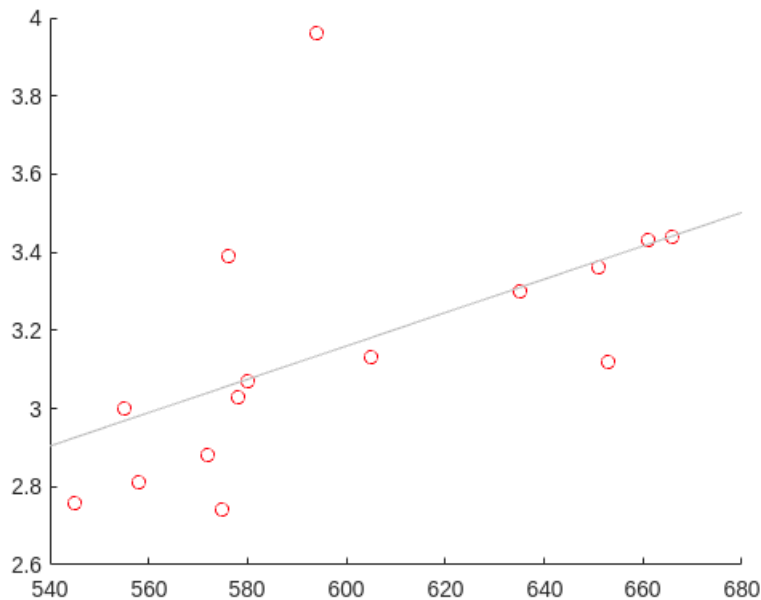


Figure 19: Línea de mínimos cuadrados entre LSAT y GPA

La estimación plug-in del coeficiente de correlación obtuvo como resultado 0.5459. Además, como se puede observar en la Figura 20, existe una correlación positiva entre LSAT Y GPA. Realizamos bootstrap para calcular la significancia estadística.

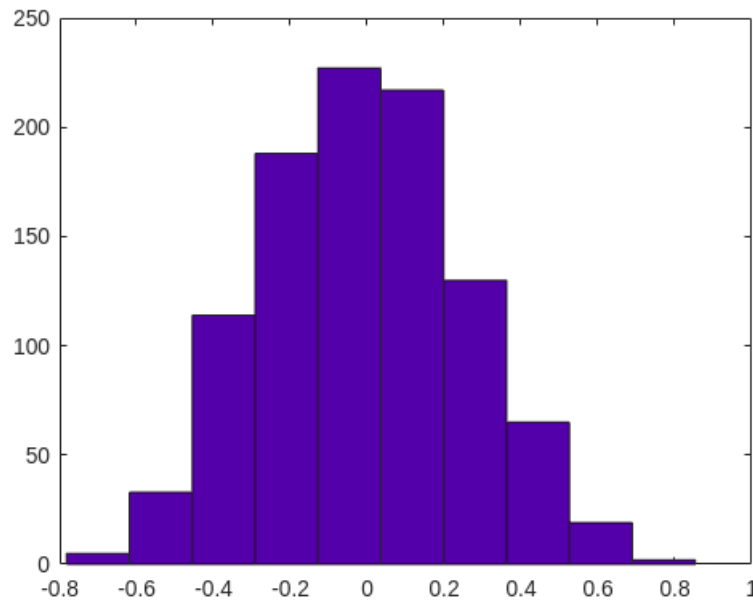


Figure 20: Histograma del bootstrap, remuestreo de 1000 para la correlación entre LSAT y GPA

Ejercicio 6 - Página 40

(Experimento computacional) Realice una simulación para comparar los cuatro métodos de intervalo de confianza bootstrap. Sea $n = 50$ y sea $T(F) = \int (x - \mu)^3 dF(x) / \sigma^3$ la asimetría. Graficar $Y_1, \dots, Y_n \sim N(0, 1)$ y establecer $X_i = e^{Y_i}, i = 1, \dots, n$. Construya los cuatro tipos de intervalos bootstrap del 95% para $T(F)$ a partir de los datos $X_1 \dots X_n$. Repita todo esto muchas veces y estime la cobertura real de los cuatro intervalos.

References

- [1] *L-statistic*. Feb. 2017. URL: <https://en.wikipedia.org/wiki/L-statistic#:~:text=In%20statistics%2C%20an%20L%2Dstatistic,statistics%20as%20estimators%20for%20parameters>.
- [2] *M-estimator*. Feb. 2017. URL: <https://en.wikipedia.org/wiki/M-estimator>.
- [3] E Ramalle-Gómara and JM Andrés De Llano. “Utilización de métodos robustos en la estadística inferencial”. In: *Atención Primaria* 32.3 (2003), p. 177.
- [4] “Robust statistics”. In: *Wikipedia* (Aug. 2022). URL: https://en.wikipedia.org/wiki/Robust_statistics#Breakdown_point.
- [5] Robert J. Serfling. “Generalized L-, M-, and R-Statistics”. In: *The Annals of Statistics* 12.1 (1984), pp. 76–86. ISSN: 00905364. URL: <http://www.jstor.org/stable/2241035> (visited on 09/19/2022).