Trabajo Práctico - IECD 2C2024

Federica di Tullio, Manuel Fernández Burda, Nicolás Letterio, Joaquín Salva

7 de diciembre de 2024

1. Resolución de problemas teóricos

Pregunta 5: Bajo $H_0 : F_X = F_Y$ (los tratamientos son indistinguibles) y asumiendo que la asignación de cada individuo al tratamiento se realiza de forma aleatoria, la distribución conjunta $F_{X,Y}$ del vector aleatorio $(\mathbf{X}, \mathbf{Y})$ es la misma que la del vector $(\mathbf{Y}, \mathbf{X})$; es decir, $F_{X,Y} = F_{Y,X}$. Probar que entonces la distribución de $\mathbf{D} = \mathbf{X} - \mathbf{Y}$ es simétrica alrededor del cero.

Sea $\mathbf{D} = \mathbf{X} - \mathbf{Y}$. Entonces la función de densidad de $\mathbf{D}$ es

$$f_D(d) = \int_{-\infty}^{+\infty} f_{X,Y}(x, x-d)\, dx$$

Ya que toma todos los valores donde $d = x - y \Leftrightarrow y = x - d$. Como $f_{X,Y}(x, y) = f_{Y,X}(x, y)$, esto implica que $f_{X,Y}(x, y) = f_{X,Y}(y, x)$. Entonces, $f_{X,Y}(x, x-d) = f_{X,Y}(x-d, x)$, por lo que:

$$f_D(d) = \int_{-\infty}^{+\infty} f_{X,Y}(x-d, x)\, dx = \int_{-\infty}^{+\infty} f_{X,Y}(u, u+d)\, du = \int_{-\infty}^{+\infty} f_{X,Y}(u, u-(-d))\, du = f_D(-d)$$

Usando el cambio de variable $u = x - d$ (y en ese caso $du = dx$) en la segunda igualdad y la definición de arriba para la última. Por lo tanto, la distribución de $\mathbf{D}$ es simétrica alrededor del 0, como queríamos demostrar.

Pregunta 6: Muestre que los siguientes estadísticos:

$$T^+ = \sum_{i=1}^{n} \mathbb{1}_{\{X_i>0\}} R_i$$

$$T^- = \sum_{i=1}^{n} \mathbb{1}_{\{X_i<0\}} R_i$$

son equivalentes a T (i.e., muestre que a partir de cualquiera de los 3 y conociendo n, se pueden computar exactamente los otros dos).
Sugerencia: calcule $T^+ + T^-$.

Para este problema vamos a usar que $T^+ + T^- = \frac{n(n+1)}{2}$ y que $T = T^+ - T^-$. Primero probémoslo.

- $T^+ + T^- = \frac{n(n+1)}{2}$

$$T^+ + T^- = \sum_{i=1}^{n} \mathbb{1}_{\{X_i>0\}} R_i + \sum_{i=1}^{n} \mathbb{1}_{\{X_i<0\}} R_i = \sum_{i=1}^{n} \left(\mathbb{1}_{\{X_i>0\}} + \mathbb{1}_{\{X_i<0\}}\right) R_i$$

Sea $i \in [n]$, como $\{X_i > 0\} \cap \{X_i < 0\} = \phi$ y $\{X_i > 0\} \cup \{X_i < 0\} = \{X_i \neq 0\}$ entonces $\mathbb{1}_{\{X_i>0\}} + \mathbb{1}_{\{X_i<0\}} = \mathbb{1}_{\{X_i\neq 0\}}$. Además $\mathbb{P}(X_i = 0) = 0$ (X_i es una variable aleatoria continua). Por lo tanto $\mathbb{1}_{\{X_i\neq 0\}} = 1$ de manera casi segura $\forall i \in [n]$.

Por lo que nos queda que

$$T^+ + T^- = \sum_{i=1}^{n} R_i$$

Ahora, $R_i \in r$ $\forall i \in [n]$ con $R_i \neq R_j$ $\forall i \neq j$. Como r es una permutación de $[n]$ y la suma es conmutativa.

$$\sum_{i=1}^{n} R_i = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

Donde la segunda igualdad se concluye de la suma de Gauss.

- $T = T^+ - T^-$

$$T^+ - T^- = \sum_{i=1}^{n} \mathbb{1}_{\{X_i>0\}} R_i - \sum_{i=1}^{n} \mathbb{1}_{\{X_i<0\}} R_i = \sum_{i=1}^{n} \left(\mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}}\right) R_i$$

Analicemos $\mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}}$. Sea $X_i \in \mathbf{X}$

  - Si $x_i < 0 \Rightarrow \mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}} = 0 - 1 = -1$
  - Si $x_i = 0 \Rightarrow \mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}} = 0 - 0 = 0$
  - Si $x_i > 0 \Rightarrow \mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}} = 1 - 0 = 1$

Entonces

$$\mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}} = \begin{cases} -1 & \text{si } x_i < 0 \\ 0 & \text{si } x_i = 0 \\ 1 & \text{si } x_i > 0 \end{cases}$$

Por lo que $\mathbb{1}_{\{X_i>0\}} - \mathbb{1}_{\{X_i<0\}} = \text{signo}(X_i)$ y entonces

$$T^+ - T^- = \sum_{i=1}^{n} \text{signo}(X_i)\, R_i = T$$

Donde la segunda igualdad es por definición.

Nos queda ver que a partir de uno de los tres y conociendo n se pueden computar los otros dos. Para hacer T en base a T^- y viceversa

$$T = T^+ - T^-$$

$$T + T^- = T^+ + T^- - T^- = \frac{n(n+1)}{2} - T^-$$

$$T = \frac{n(n+1)}{2} - 2T^-$$

Para hacer T en base a T^+ y viceversa

$$T = T^+ - T^-$$

$$T - T^+ = T^+ - (T^- + T^+) = T^+ - \frac{n(n+1)}{2}$$

$$T = 2T^+ - \frac{n(n+1)}{2}$$

Y concluimos que

$$T = 2T^+ - \frac{n(n+1)}{2} = \frac{n(n+1)}{2} - 2T^-$$

como queríamos ver.

Pregunta 7: Demuestre que bajo H_0, $|X_i|$ es independiente de $\text{signo}(X_i)$.
Sugerencia: utilice sus conocimientos sobre F_X bajo H_0.

Para ver que son independientes es necesario que

$$\mathbb{P}(\text{signo}(X_i) \leq s, |X_i| \leq t) = \mathbb{P}(\text{signo}(X_i) \leq s)\, \mathbb{P}(|X_i| \leq t) \quad \forall s \in \{-1, 0, 1\}, t \in \mathbb{R}$$

Como $Im(\text{signo}) = \{-1, 0, 1\}$ es suficiente con ver que

$$\mathbb{P}(\text{signo}(X_i) = k, |X_i| \leq t) = \mathbb{P}(\text{signo}(X_i) = k)\, \mathbb{P}(|X_i| \leq t) \quad \forall k \in \{-1, 0, 1\}, t \in \mathbb{R}$$

Hagamos los casos de k. Sea $t \in \mathbb{R}$.

- Si $k = 0$

  Como X_i es una variable aleatoria continua

  $$\mathbb{P}(\text{signo}(X_i) = 0, |X_i| \leq t) = \mathbb{P}(|X_i| = 0) = 0$$

  Y por lo mismo

  $$\mathbb{P}(\text{signo}(X_i) = 0) = \mathbb{P}(X_i = 0) = 0$$

  Entonces

  $$\mathbb{P}(\text{signo}(X_i) = 0, |X_i| \leq t) = 0 = \mathbb{P}(\text{signo}(X_i) = 0)\, \mathbb{P}(|X_i| \leq t)$$

- Si $k = 1$

  Abriendo el modulo sabemos que

  $$\mathbb{P}(\text{signo}(X_i) = 1, |X_i| \leq t) = \mathbb{P}(\text{signo}(X_i) = 1, -t \leq X_i \leq t) = \mathbb{P}(0 \leq X_i \leq t) = \frac{1}{2}\, 2\, \mathbb{P}(0 \leq X_i \leq t)$$

  Como bajo H_0 sabemos que F_{X_i} es simétrica alrededor del 0, $\mathbb{P}(0 \leq X_i \leq t) = \mathbb{P}(-t \leq X_i \leq 0)$ entonces

  $$2\, \mathbb{P}(0 \leq X_i \leq t) = \mathbb{P}(0 \leq X_i \leq t) + \mathbb{P}(-t \leq X_i \leq 0) = \mathbb{P}(-t \leq X_i \leq t) = \mathbb{P}(|X_i| \leq t)$$

  Por otra parte $\mathbb{1}_{\{\text{signo}(X_i)=1\}} = \mathbb{1}_{\{X_i > 0\}} \sim Be(\frac{1}{2})$ como veremos más tarde en la **Pregunta 9**. Por lo que

  $$\mathbb{P}(\text{signo}(X_i) = 1) = \mathbb{E}\left(\mathbb{1}_{\{X_i > 0\}}\right) = \frac{1}{2}$$

Y concluimos que

$$\mathbb{P}(\text{signo}(X_i) = 1, |X_i| \leq t) = \frac{1}{2}\, 2\, \mathbb{P}(0 \leq X_i \leq t) = \mathbb{P}(\text{signo}(X_i) = 1)\, \mathbb{P}(|X_i| \leq t)$$

Como queríamos ver.

- Si $k = -1$

  Al igual que en el caso anterior $\mathbb{1}_{\{\text{signo}(X_i)=-1\}} = \mathbb{1}_{\{X_i<0\}} \sim Be\left(\frac{1}{2}\right)$ y $2\,\mathbb{P}(-t \leq X_i \leq 0) = \mathbb{P}(|X_i| \leq t)$. Por lo que simétricamente

$$\mathbb{P}(\text{signo}(X_i) = 1, |X_i| \leq t) = \mathbb{P}(-t \leq X_i \leq 0) = \mathbb{P}(\text{signo}(X_i) = -1)\, \mathbb{P}(|X_i| \leq t)$$

Y queda probado que $\text{signo}(X_i)$ es independiente de $|X_i|$.

Pregunta 8: A partir del resultado anterior, pruebe que bajo H_0 los vectores de rangos $\mathbf{R} = (R_1, \ldots, R_n)$ y antirrangos $\mathbf{D} = (D_1, \ldots, D_n)$ correspondientes a $|\mathbf{X}|$ son independientes del vector de signos $\mathbf{S} = (\text{signo}(X_1), \ldots, \text{signo}(X_n))$ de la muestra original $\mathbf{X}$.

Sea $i \in [n]$. Como $\text{signo}(X_i)$ es independiente de $|X_i|$ como vimos en la **Pregunta 7** y tanto R_i como D_i son funciones que dependen exclusivamente de $|X_i|$ por lo que también son independientes de $\text{signo}(X_i)$. Como i es genérico, esto para para todo $i \in [n]$.

Por otra parte, como $X_1, \ldots, X_n$ son independientes entre ellos dos a dos resulta que para $i \neq j \in [n]$, R_i, R_j van a ser independientes al igual que D_i, D_j y, $\text{signo}(X_i)$, $\text{signo}(X_j)$. Entonces $\mathbf{R}$ y $\mathbf{D}$ son independientes de $\mathbf{S}$.

Pregunta 9: Pruebe que bajo $H_0 : \theta = 0$, $F \in \Omega_s$, las v.a. $W_j = \mathbb{1}_{\{X_{D_j}>0\}}$ se distribuyen según

$$W_1, \ldots, W_n \overset{\text{iid}}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right)$$

Sugerencia: Utilice la Ley de la Probabilidad Total de $\boldsymbol{W}$ sobre los posibles valores de $\boldsymbol{D}$ y utilice los resultados previos para operar y factorizar $\mathbb{P}(\mathbf{W} = \mathbf{w}) = \prod_{j\in[n]} \mathbb{P}(W_j = w_j)$

Veamos primero que se distribuyen como $Be\left(\frac{1}{2}\right)$. Sea $j \in [n]$, sabemos que

$$W_j = \mathbb{1}_{\{X_{D_j}>0\}} = \mathbb{1}_{\{X_i>0\}} \quad \text{para algún } i \in [n]$$

Observemos que si probamos que $\mathbb{P}(\mathbb{1}_{\{X_i>0\}} = 1) = \frac{1}{2}$ es suficiente pues el complemento es que sea igual a 0.

Estando bajo H_0 sabemos que la densidad de X_i es simétrica. Entonces $\int_0^{+\infty} f_{X_i}(x)\,dx = \int_{-\infty}^0 f_{X_i}(x)\,dx$. De esto podemos desarrollar que

$$\mathbb{P}(\mathbb{1}_{\{X_i>0\}} = 1) = \mathbb{P}(X_i > 0) = \int_0^{+\infty} f_{X_i}(x)\,dx = \frac{1}{2}\, 2 \int_0^{+\infty} f_{X_i}(x)\,dx$$

$$= \frac{1}{2}\left(\int_0^{+\infty} f_{X_i}(x)\,dx + \int_{-\infty}^0 f_{X_i}(x)\,dx\right) = \frac{1}{2}\int_{-\infty}^{+\infty} f_{X_i}(x)\,dx = \frac{1}{2}$$

Donde en la última igualdad usamos que como f_{X_i} es una densidad, su integral en todo $\mathbb{R}$ es igual a 1. Por lo que W_i se distribuye como $Be\left(\frac{1}{2}\right)$ $\forall i \in [n]$ que era lo que queríamos probar.

Veamos ahora que los W_j son independientes entre ellos. Para ello es necesario que

$$\mathbb{P}(\boldsymbol{w} = \boldsymbol{a}) = \prod_{j=1}^{n} \mathbb{P}(w_j = a_j) \quad \forall \boldsymbol{a} \in \{1, 0\}^n$$

Sea $\boldsymbol{a} \in \{1,0\}^n$. Observemos que $\boldsymbol{d} \in S_n = \{(d_1,\dots,d_n)|1 \leq d_i \neq d_j \leq n\}$ o sea, las permutaciones de $[n]$. Usando la ley de probabilidad total podemos desarrollar

$$\mathbb{P}(\mathbf{W}=\boldsymbol{a}) = \sum_{\boldsymbol{d}\in S_n} \mathbb{P}(\boldsymbol{w}=\boldsymbol{a}|\mathbf{D}=\boldsymbol{d}) = \sum_{\boldsymbol{d}\in S_n} \mathbb{P}\left(\mathbb{1}_{\{X_{d_1}>0\}}=a_1,\dots,\mathbb{1}_{\{X_{d_n}>0\}}=a_n|\mathbf{D}=\boldsymbol{d}\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d})$$

Como sabemos que bajo H_0 $\mathbf{S}$ y $\mathbf{D}$ son independientes podemos dejar de condicionar a $\mathbf{D}=\boldsymbol{d}$. Lo de arriba entonces nos queda que

$$\mathbb{P}(\mathbf{W}=\boldsymbol{a}) = \sum_{\boldsymbol{d}\in S_n} \mathbb{P}\left(\mathbb{1}_{\{X_{d_1}>0\}}=a_1,\dots,\mathbb{1}_{\{X_{d_n}>0\}}=a_n\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d})$$

Además signo(X_i) y signo(X_j) son independientes $\forall i \neq j \in [n]$ entonces

$$\sum_{\boldsymbol{d}\in S_n} \mathbb{P}\left(\mathbb{1}_{\{X_{d_1}>0\}}=a_1,\dots,\mathbb{1}_{\{X_{d_n}>0\}}=a_n\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d}) = \sum_{\boldsymbol{d}\in S_n}\left(\prod_{j=1}^n \mathbb{P}\left(\mathbb{1}_{\{X_{d_j}>0\}}=a_j\right)\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d})$$

Usemos que si $j\in[n]$, $\mathbb{P}\left(\mathbb{1}_{\{X_{d_j}>0\}}=a_j\right)=\frac{1}{2}$ pues $\mathbb{1}_{\{X_{d_j}>0\}} \sim Be\left(\frac{1}{2}\right)$. Por lo que

$$\sum_{\boldsymbol{d}\in S_n}\left(\prod_{j=1}^n \mathbb{P}(\mathbb{1}_{\{X_{d_j}>0\}}=a_j)\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d}) = \sum_{\boldsymbol{d}\in S_n}\left(\prod_{j=1}^n \frac{1}{2}\right)\mathbb{P}(\mathbf{D}=\boldsymbol{d}) =$$

$$= \sum_{\boldsymbol{d}\in S_n}\left(\frac{1}{2}\right)^n \mathbb{P}(\mathbf{D}=\boldsymbol{d}) = \left(\frac{1}{2}\right)^n \sum_{\boldsymbol{d}\in S_n}\mathbb{P}(\mathbf{D}=\boldsymbol{d})$$

Por ley de probabilidad total, como todos los posibles $\mathbf{D}$ son todos los elementos de S_n, entonces $\sum_{\boldsymbol{d}\in S_n}\mathbb{P}(\mathbf{D}=\boldsymbol{d})=1$. Por lo tanto

$$\mathbb{P}(\mathbf{W}=\boldsymbol{a}) = \left(\frac{1}{2}\right)^n \sum_{\boldsymbol{d}\in S_n}\mathbb{P}(\mathbf{D}=\boldsymbol{d}) = \left(\frac{1}{2}\right)^n = \prod_{j=1}^n \mathbb{P}(\mathbb{1}_{\{X_{d_j}>0\}}=a_j) = \prod_{j=1}^n \mathbb{P}(w_j=a_j)$$

Como queríamos ver.

Pregunta 11: Muestre que T^+ es simétrica alrededor de $\frac{n(n+1)}{4}$

Por definición sabemos que T^+ es simétrica alrededor de $\frac{n(n+1)}{4}$ sii $T^+ - \frac{n(n+1)}{4} \sim -(T^+ - \frac{n(n+1)}{4})$ que es lo mismo que $F_{T^+-\frac{n(n+1)}{4}}(t) = F_{-T^++\frac{n(n+1)}{4}}(t), \forall t\in\mathbb{R}$. Veamos que esto es cierto.

Primero desarrollemos las acumuladas. Como T^+ es una variable aleatoria discreta con $rg(T^+) = \{0,1,\dots,\frac{n(n+1)}{2}\}$ la probabilidad de que T^+ sea igual a algo fuera de este rango es 0, entonces sin pérdida de generalidad tomemos que $t+\frac{n(n+1)}{4}, -t+\frac{n(n+1)}{4} \in rg(T^+)$.

Entonces si analizamos la acumulada de la izquierda tenemos que

$$F_{T^+-\frac{n(n+1)}{4}}(t) = \mathbb{P}\left(T^+ - \frac{n(n+1)}{4} \leq t\right) = \mathbb{P}\left(T^+ \leq t + \frac{n(n+1)}{4}\right) = \sum_{i=0}^{t+\frac{n(n+1)}{4}} p_{T^+}(i)$$

Mientras que la acumulada de la derecha nos queda

$$F_{-T^++\frac{n(n+1)}{4}}(t) = \mathbb{P}\left(-T^+ + \frac{n(n+1)}{4} \leq t\right) = \mathbb{P}\left(-T^+ \leq t - \frac{n(n+1)}{4}\right)$$

$$= \mathbb{P}\left(T^+ \geq -t + \frac{n(n+1)}{4}\right) = \sum_{i=-t+\frac{n(n+1)}{4}}^{\frac{n(n+1)}{2}} p_{T^+}(i)$$

Si vemos que $p_{T^+}(i) = p_{T^+}\left(\frac{n(n+1)}{2} - i\right)$ es suficiente pues con el cambio de variable $i = \frac{n(n+1)}{2} - j$ en la última sumatoria como $-t + \frac{n(n+1)}{4} \leq \frac{n(n+1)}{2} - j \leq \frac{n(n+1)}{2} \Leftrightarrow t + \frac{n(n+1)}{4} \geq j \geq 0$ nos queda que

$$\sum_{i=-t+\frac{n(n+1)}{4}}^{\frac{n(n+1)}{2}} p_{T^+}(i) = \sum_{j=0}^{t+\frac{n(n+1)}{4}} p_{T^+}\left(\frac{n(n+1)}{2} - j\right)$$

Entonces pasaría que

$$F_{T^+ - \frac{n(n+1)}{4}}(t) = \sum_{j=0}^{t+\frac{n(n+1)}{4}} p_{T^+}(i) = \sum_{j=0}^{t+\frac{n(n+1)}{4}} p_{T^+}\left(\frac{n(n+1)}{2} - j\right) = F_{-T^+ + \frac{n(n+1)}{4}}(t)$$

Como queríamos ver. Probemos entonces que $p_{T^+}(i) = p_{T^+}(\frac{n(n+1)}{2} - i)$, es decir que $\mathbb{P}(T^+ = i) = \mathbb{P}(T^+ = \frac{n(n+1)}{2} - i)$. Usando los resultados de la **Pregunta 6** sabemos que

$$\mathbb{P}\left(T^+ = \frac{n(n+1)}{2} - i\right) = \mathbb{P}\left(T^+ - (T^+ + T^-) = -i\right) = \mathbb{P}(-T^- = -i) = \mathbb{P}(T^- = i)$$

Pero $\mathbb{P}(T^- = i) = \mathbb{P}(T^+ = i)$ porque $T^- \sim T^+$ pues

$$\mathbb{P}(X_i < 0) = \mathbb{P}(X_i > 0), \forall i \in [n] \Rightarrow \mathbb{1}_{\{X_i<0\}} \sim \mathbb{1}_{\{X_i>0\}}, \forall i \in [n] \Rightarrow \sum_{i=1}^{n} \mathbb{1}_{\{X_i<0\}} R_i \sim \sum_{i=1}^{n} \mathbb{1}_{\{X_i>0\}} R_i$$

Pues como las X_i son independientes entre ellas, las indicadoras entre ellas también y entonces sumarlas y multiplicarles lo mismo también queda idénticamente distribuido. Y estas dos ultimas distribuciones son las de T^- y T^+ respectivamente.

Pregunta 15: Bajo H_0, ¿cuánto vale $\mathbb{E}(T^+)$? ¿Y $\text{Var}(T^+)$?

- $\mathbb{E}(T^+)$

  Por definición $T^+ = \sum_{j=1}^{n} \mathbb{1}_{\{X_{D_j}>0\}} j$ entonces, por linealidad de la esperanza

  $$\mathbb{E}(T^+) = \mathbb{E}\left(\sum_{j=1}^{n} \mathbb{1}_{\{X_{D_j}>0\}} j\right) = \sum_{j=1}^{n} j\, \mathbb{E}\left(\mathbb{1}_{\{X_{D_j}>0\}}\right)$$

  Pero bajo H_0 como X_{D_j} es simétrica alrededor del 0 implica que $\mathbb{E}(\mathbb{1}_{\{X_{D_j}>0\}}) = \mathbb{P}(X_{D_j} > 0) = \frac{1}{2}$. Por lo que

  $$\sum_{j=1}^{n} j\, \mathbb{E}\left(\mathbb{1}_{\{X_{D_j}>0\}}\right) = \sum_{j=1}^{n} j\, \frac{1}{2} = \frac{1}{2} \sum_{j=1}^{n} j = \frac{n(n+1)}{4}$$

  Donde la última igualdad la concluimos de la suma de Gauss. Por lo tanto

  $$\mathbb{E}(T^+) = \frac{n(n+1)}{4}$$

- $\mathrm{Var}(T^+)$

  Al estar bajo H_0, sabemos que $D_1, \ldots, D_n$ son independientes entre sí, entonces $\mathbb{1}_{\{X_{D_1}>0\}}\,1, \ldots, \mathbb{1}_{\{X_{D_n}>0\}}\,n$ también lo son entre ellas. Así que la varianza se distribuye en la sumatoria.

  $$\mathrm{Var}(T^+) = \mathrm{Var}\left(\sum_{j=1}^{n} \mathbb{1}_{\{X_{D_j}>0\}}\, j\right) = \sum_{j=1}^{n} \mathrm{Var}\left(\mathbb{1}_{\{X_{D_j}>0\}}\, j\right) = \sum_{j=1}^{n} j^2\, \mathrm{Var}\left(\mathbb{1}_{\{X_{D_j}>0\}}\right)$$

  En la última igualdad usamos que la varianza saca constantes al cuadrado. Por definición de indicadora y resultados de probabilidad $\mathrm{Var}\left(\mathbb{1}_{\{X_{D_j}>0\}}\right) = \mathbb{P}(X_{D_j} > 0)\,\mathbb{P}(X_{D_j} \leq 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Por lo tanto

  $$\mathrm{Var}(T^+) = \sum_{j=1}^{n} j^2\, \mathrm{Var}\left(\mathbb{1}_{\{X_{D_j}>0\}}\right) = \sum_{j=1}^{n} \frac{1}{4}\, j^2 = \frac{n(n+1)(2n+1)}{24}$$

  Que es lo que queríamos calcular, usando que la última sumatoria también tiene fórmula cerrada como la de Gauss.

Pregunta 16: Dé la distribución asintótica de T^+.

Para este problema usemos el TCL de Lindeberg. Tomemos para cada $i \in [n]$, $W_i = \mathbb{1}_{\{X_{D_i}>0\}} - \frac{1}{2}$ para que cumpla con las hipótesis necesarias. Ya que son i.i.d., porque las X_i lo son, y $\mathbb{E}(\mathbb{1}_{\{X_{D_i}\}} - \frac{1}{2}) = 0$ pues

$$\mathbb{E}\left(\mathbb{1}_{\{X_{D_i}>0\}} - \frac{1}{2}\right) = \mathbb{E}\left(\mathbb{1}_{\{X_{D_i}>0\}}\right) - \frac{1}{2} = \mathbb{P}(X_{D_i>0}) - \frac{1}{2} = \frac{1}{2} - \frac{1}{2} = 0$$

donde la primer igualdad es porque la esperanza es lineal, la segunda por propiedades de la indicadora y la tercera porque X_{D_i} es simétrica alrededor del 0.

Entonces definamos la S del teorema como

$$S = \sum_{i=1}^{n} \frac{i\left(\mathbb{1}_{\{X_{D_i}>0\}} - \frac{1}{2}\right)}{\sqrt{n}}$$

Veamos que $\frac{\max_i |i|}{\sqrt{\sum_{i=1}^n i^2}} \underset{n\to\infty}{\longrightarrow} 0$ para terminar de cumplir con las hipótesis del teorema.

$$\frac{\max\limits_i |i|}{\sqrt{\sum_{i=1}^n i^2}} = \frac{n}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} = \sqrt{\frac{6n}{(n+1)(2n+1)}} \underset{n\to\infty}{\longrightarrow} 0$$

La primer igualdad la inducimos de la fórmula cerrada de $\sum_{i=1}^n i^2$ y la convergencia de análisis del CBC.

Calculemos ahora la varianza de S. Para eso veamos cual es la de W_1 primero.

$$\mathrm{Var}(W_1) = \mathrm{Var}\left(\mathbb{1}_{\{X_{D_1}>0\}} - \frac{1}{2}\right) = \mathrm{Var}\left(\mathbb{1}_{\{X_{D_1}>0\}}\right) = \mathbb{P}(X_{D_1} > 0)\,\mathbb{P}(X_{D_1} \leq 0) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

Pues la varianza anula suma de constantes y por un resultado de probabilidad sobre la indicadora en la tercer igualdad. Entonces

$$\mathrm{Var}(S) = \frac{1}{4}\left(\sum_{i=1}^{n} i^2\right)\frac{1}{n} = \frac{\cancel{n}(n+1)(2n+1)}{\cancel{n}24}$$

Por el teorema sabemos que $\frac{S}{\sqrt{\operatorname{Var}(S)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. Desarrollemoslo para ver la distribución asintótica de T^{+}.

$$\frac{S}{\sqrt{\operatorname{Var}(S)}} = \frac{\sum_{i=1}^{n} \frac{i\left(\mathbb{1}_{\{X_{D_i}>0\}}-\frac{1}{2}\right)}{\sqrt{n}}}{\sqrt{\frac{(n+1)(2n+1)}{24}}} = \frac{\sqrt{24}\left(\left(\sum_{i=1}^{n} i\mathbb{1}_{\{X_{D_i}\}}\right)-\frac{1}{2}\sum_{i=1}^{n} i\right)}{\sqrt{n(n+1)(2n+1)}} = \frac{\sqrt{24}\left(T^{+}-\frac{n(n+1)}{4}\right)}{\sqrt{n(n+1)(2n+1)}}$$

Por lo tanto, la distribución asintótica de T^{+} nos queda

$$\frac{\sqrt{24}\left(T^{+}-\frac{n(n+1)}{4}\right)}{\sqrt{n(n+1)(2n+1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Que tiene sentido porque

$$\frac{\sqrt{24}\left(T^{+}-\frac{n(n+1)}{4}\right)}{\sqrt{n(n+1)(2n+1)}} = \frac{T^{+}-\mathbb{E}(T^{+})}{\sqrt{\operatorname{Var}(T^{+})}}$$

2. Gráficos

Pregunta 17: Fije $n_1 = 4, n_2 = 10, n_3 = 20$. Para cada n, realice un gráfico de barras con la probabilidad puntual *exacta* de T^{+} (válgase de `dTmas`) y superpóngale una línea con la densidad asintótica esperada. ¿Coinciden razonablemente? ¿En toda la distribución, en el centro, en las colas? ¿A partir de qué n? ¿Se le ocurre alguna corrección sencilla para los n pequeños?

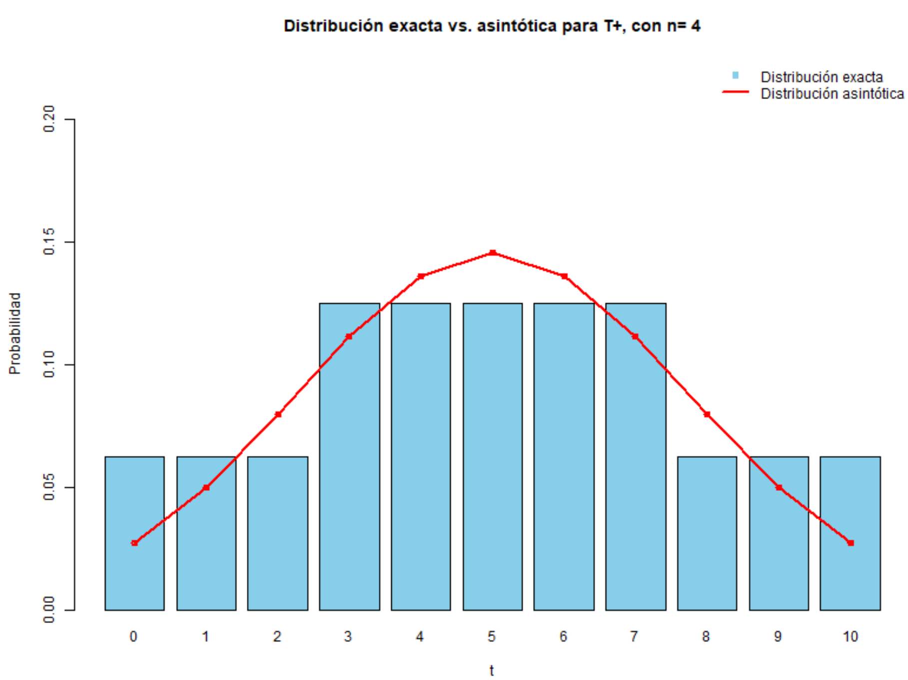

Figura 1: Fijando $n_1 = 4$

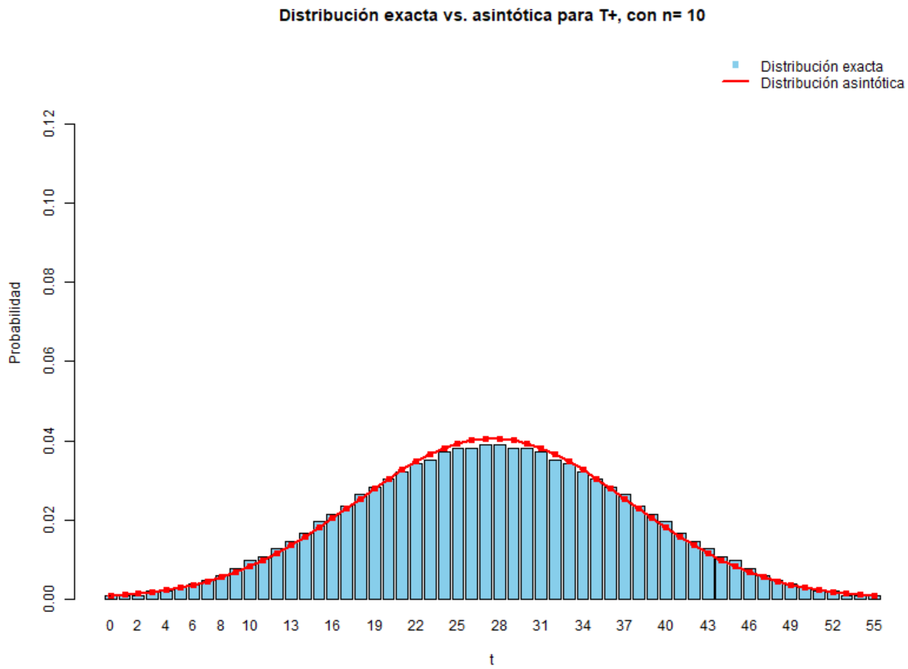

Figura 2: Fijando $n_2 = 10$

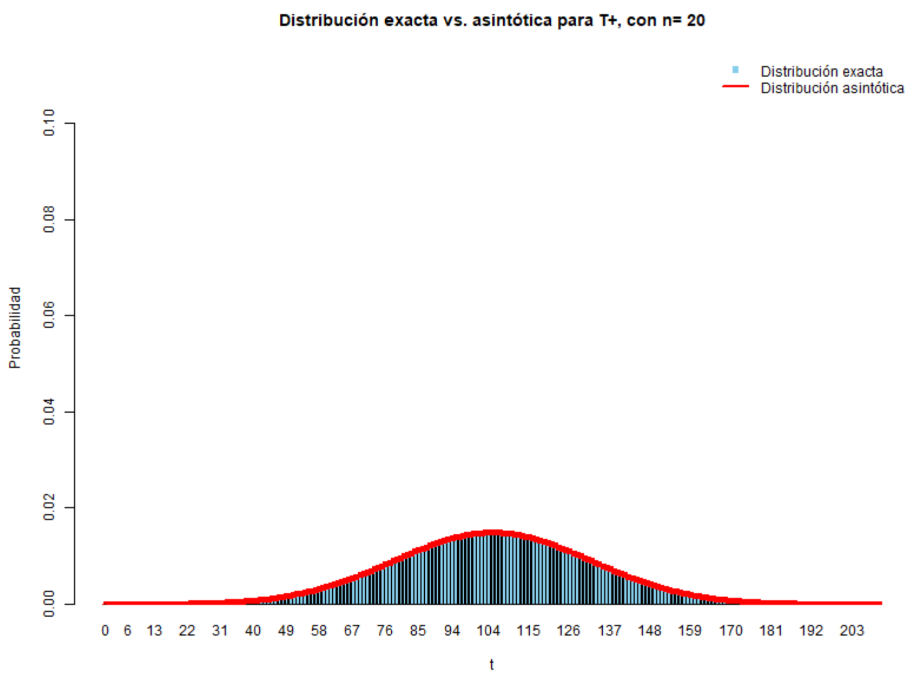

Figura 3: Fijando $n_3 = 20$

Podemos ver que coinciden razonablemente a partir de $n = 10$. Coinciden para todos los n en el centro y en las colas. Tal vez, para $n = 4$, la coincidencia en las colas es cuestionable, pero el comportamiento es similar (densidad concentrada en el centro y simetría respecto de éste).