

Distribución asintótica del estadígrafo T^+ de los rangos signados de Wilcoxon

claude.ai

14 de abril de 2026

Índice

1. Planteamiento y notación	1
2. Independencia rangos–signos y representación clave	2
3. Esperanza y varianza de T^+	3
4. Normalidad asintótica: Teorema Central del Límite de Lindeberg	3
4.1. El TCL de Lindeberg	3
4.2. Aplicación a T^+	3
5. Resumen y uso práctico	5

1. Planteamiento y notación

Sea X_1, \dots, X_n una muestra aleatoria simple procedente de una distribución continua F simétrica respecto a la mediana θ_0 . El contraste de hipótesis es

$$H_0 : \theta_0 = 0 \quad \text{frente a} \quad H_1 : \theta_0 \neq 0.$$

Definición 1.1 (Rangos y estadígrafo T^+). Bajo H_0 , se trabaja con las observaciones X_1, \dots, X_n directamente (ya centradas en cero).

1. Sea R_i el *rango* de $|X_i|$ entre $|X_1|, \dots, |X_n|$ (se supone que no hay empates c.s., lo cual se garantiza por la continuidad de F).
2. Sea $\varepsilon_i = \text{sgn}(X_i) \in \{-1, +1\}$.

3. El estadígrafo de Wilcoxon de rangos signados es

$$T^+ := \sum_{i=1}^n R_i \mathbf{1}(\varepsilon_i = +1) = \sum_{i=1}^n R_i \mathbf{1}(X_i > 0).$$

2. Independencia rangos–signos y representación clave

El resultado estructural fundamental es el siguiente.

Lema 2.1 (Independencia rangos–signos bajo H_0). *Si F es continua y simétrica respecto a 0, entonces:*

(a) *El vector de rangos (R_1, \dots, R_n) y el vector de signos $(\varepsilon_1, \dots, \varepsilon_n)$ son mutuamente independientes.*

(b) *Los signos $\varepsilon_1, \dots, \varepsilon_n$ son variables aleatorias i.i.d. con $\mathbb{P}(\varepsilon_i = +1) = \frac{1}{2}$.*

Demostración (esquema). Como F es simétrica en 0, para cada i los eventos $\{X_i > 0\}$ y $\{X_i < 0\}$ tienen probabilidad $\frac{1}{2}$ y son independientes del valor absoluto $|X_i|$, pues la función de densidad satisface $f(x) = f(-x)$. Más formalmente, la transformación $(X_i) \mapsto (|X_i|, \text{sgn}(X_i))$ es bimeasurable y la medida imagen se factoriza como el producto de la distribución de $|X_i|$ (que determina los rangos) y la distribución de $\text{sgn}(X_i)$ (uniforme en $\{-1, +1\}$). La simetría implica que esta factorización se mantiene para el vector completo, de modo que el σ -álgebra generada por los rangos y la generada por los signos son independientes. \square

Observación 2.2. El lema garantiza que los rangos son siempre la colección $\{1, \dots, n\}$ (solo se redistribuyen entre los i 's) y que los signos actúan como «máscaras» aleatorias independientes.

Proposición 2.3 (Representación por sumas independientes). *Bajo H_0 , el estadígrafo T^+ tiene la misma distribución que*

$$T^+ \stackrel{d}{=} \sum_{j=1}^n j Z_j,$$

donde $Z_1, \dots, Z_n \stackrel{iid}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right)$ son independientes entre sí.

Demostración. Por el Lema 2.1(a), dado que (R_1, \dots, R_n) es una permutación aleatoria uniforme de $\{1, \dots, n\}$ independiente de los signos, el vector $(R_{\sigma(1)}, \dots, R_{\sigma(n)})$ tiene la misma distribución para cualquier permutación σ inducida por los signos. En consecuencia,

$$T^+ = \sum_{i=1}^n R_i \mathbf{1}(\varepsilon_i = +1) \stackrel{d}{=} \sum_{j=1}^n j \mathbf{1}(\varepsilon_{\pi(j)} = +1),$$

donde π es la permutación que ordena los valores absolutos. Definiendo $Z_j := \mathbf{1}(\varepsilon_{\pi(j)} = +1)$, la independencia de rangos y signos del Lema 2.1 implica que los Z_j son i.i.d. Bernoulli($\frac{1}{2}$) independientes de los rangos; los pesos j pasan a ser constantes deterministas, con lo que se obtiene la representación enunciada. \square

3. Esperanza y varianza de T^+

Proposición 3.1 (Momentos exactos de T^+). *Bajo H_0 ,*

$$\mathbb{E}[T^+] = \frac{n(n+1)}{4}, \quad \text{Var}(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

Demostración. Usando la representación de la Proposición 2.3 y $\mathbb{E}[Z_j] = \frac{1}{2}$, $\text{Var}(Z_j) = \frac{1}{4}$:

Esperanza.

$$\mathbb{E}[T^+] = \sum_{j=1}^n j \mathbb{E}[Z_j] = \frac{1}{2} \sum_{j=1}^n j = \frac{1}{2} \cdot \frac{n(n+1)}{2} = \frac{n(n+1)}{4}.$$

Varianza. Como los Z_j son *independientes*, no hay términos de covarianza cruzada:

$$\text{Var}(T^+) = \sum_{j=1}^n j^2 \text{Var}(Z_j) = \frac{1}{4} \sum_{j=1}^n j^2 = \frac{1}{4} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)(2n+1)}{24}.$$

En ambos casos se ha utilizado la independencia mutua de los Z_j , consecuencia directa del Lema 2.1. \square

4. Normalidad asintótica: Teorema Central del Límite de Lindeberg

4.1. El TCL de Lindeberg

Recordamos la versión clásica del teorema que utilizaremos.

Teorema 4.1 (TCL de Lindeberg, 1922). *Sea $\{X_{n,j} : 1 \leq j \leq k_n\}_{n \geq 1}$ un array de variables aleatorias reales con $X_{n,1}, \dots, X_{n,k_n}$ independientes para cada n , $\mathbb{E}[X_{n,j}] = 0$ y $\sigma_{n,j}^2 := \text{Var}(X_{n,j}) < \infty$. Denótese $s_n^2 := \sum_{j=1}^{k_n} \sigma_{n,j}^2 > 0$. Si se cumple la **condición de Lindeberg**: para todo $\varepsilon > 0$,*

$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{j=1}^{k_n} \mathbb{E} \left[X_{n,j}^2 \mathbf{1}(|X_{n,j}| > \varepsilon s_n) \right] \xrightarrow{n \rightarrow \infty} 0,$$

entonces

$$\frac{1}{s_n} \sum_{j=1}^{k_n} X_{n,j} \xrightarrow{d} \mathcal{N}(0, 1).$$

4.2. Aplicación a T^+

Definamos el array de sumandos centrados.

Definición 4.2. Para $1 \leq j \leq n$, sea

$$X_{n,j} := j \left(Z_j - \frac{1}{2} \right),$$

de modo que $\mathbb{E}[X_{n,j}] = 0$ y

$$\sigma_{n,j}^2 = j^2 \operatorname{Var} \left(Z_j - \frac{1}{2} \right) = j^2 \cdot \frac{1}{4} = \frac{j^2}{4}.$$

La suma $S_n := \sum_{j=1}^n X_{n,j}$ satisface

$$S_n = T^+ - \mathbb{E}[T^+], \quad s_n^2 = \sum_{j=1}^n \frac{j^2}{4} = \frac{n(n+1)(2n+1)}{24} = \operatorname{Var}(T^+).$$

Teorema 4.3 (Normalidad asintótica de T^+). *Bajo H_0 ,*

$$W_n := \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (n \rightarrow \infty).$$

Demostración. Por la Proposición 2.3, $T^+ \stackrel{d}{=} \sum_{j=1}^n j Z_j$ con Z_j i.i.d. Bernoulli($\frac{1}{2}$) independientes. El array $\{X_{n,j}\}$ definido arriba satisface las condiciones de media cero e independencia del Teorema 4.1 con $k_n = n$. Solo resta verificar la condición de Lindeberg.

Paso 1: acotación de cada sumando.

Como $Z_j \in \{0, 1\}$, tenemos $Z_j - \frac{1}{2} \in \{-\frac{1}{2}, +\frac{1}{2}\}$, luego

$$|X_{n,j}| = j \left| Z_j - \frac{1}{2} \right| = \frac{j}{2} \leq \frac{n}{2} \quad \forall j = 1, \dots, n.$$

Paso 2: comportamiento asintótico de s_n .

$$s_n^2 = \frac{n(n+1)(2n+1)}{24} \sim \frac{n^3}{12} \implies s_n \sim \frac{n^{3/2}}{2\sqrt{3}} \quad (n \rightarrow \infty).$$

En particular,

$$\frac{n/2}{s_n} \sim \frac{n/2}{n^{3/2}/(2\sqrt{3})} = \frac{\sqrt{3}}{n^{1/2}} \xrightarrow{n \rightarrow \infty} 0.$$

Paso 3: verificación de la condición de Lindeberg.

Sea $\varepsilon > 0$ arbitrario. Por el Paso 2, existe $n_0 = n_0(\varepsilon) \in \mathbb{N}$ tal que, para todo $n \geq n_0$,

$$\frac{n}{2} < \varepsilon s_n,$$

es decir, $|X_{n,j}| \leq \frac{n}{2} < \varepsilon s_n$ para todo $j \in \{1, \dots, n\}$. Por lo tanto, el evento $\{|X_{n,j}| > \varepsilon s_n\}$ tiene probabilidad cero para cada j , y en consecuencia:

$$L_n(\varepsilon) = \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}[X_{n,j}^2 \mathbf{1}(|X_{n,j}| > \varepsilon s_n)] = 0 \quad \forall n \geq n_0.$$

La condición de Lindeberg queda verificada (de hecho se cumple *exactamente a partir de* n_0 , no solo en el límite).

Conclusión.

El Teorema 4.1 garantiza que

$$W_n = \frac{S_n}{s_n} = \frac{T^+ - \mathbb{E}[T^+]}{\sqrt{\text{Var}(T^+)}} \xrightarrow{d} \mathcal{N}(0, 1),$$

lo cual completa la demostración. □

Observación 4.4 (Relación con la condición de Lyapunov). La verificación del Paso 3 es más fuerte que la condición de Lyapunov de orden $2 + \delta$ con $\delta = 1$, que exigiría $\sum_j \mathbb{E}[|X_{n,j}|^3]/s_n^3 \rightarrow 0$. En efecto, como cada $|X_{n,j}| \leq n/2$, se tiene

$$\frac{1}{s_n^3} \sum_{j=1}^n \mathbb{E}[|X_{n,j}|^3] \leq \frac{n/2}{s_n^3} \sum_{j=1}^n \mathbb{E}[X_{n,j}^2] = \frac{n/2}{s_n^3} \cdot s_n^2 = \frac{n}{2 s_n} \sim \frac{\sqrt{3}}{n^{1/2}} \rightarrow 0.$$

Ambos enfoques coinciden en el resultado; la verificación directa de Lindeberg es la más transparente en este caso.

5. Resumen y uso práctico

El Teorema 4.3 justifica el siguiente procedimiento asintótico.

Corolario 5.1 (Contraste asintótico). *Para n grande, bajo $H_0 : \theta_0 = 0$ y con nivel de significación $\alpha \in (0, 1)$, la región de rechazo bilateral es*

$$|W_n| > z_{\alpha/2},$$

donde $z_{\alpha/2}$ es el cuantil $(1 - \alpha/2)$ de la distribución $\mathcal{N}(0, 1)$. El valor- p aproximado del contraste es

$$p = 2\mathbb{P}(\mathcal{N}(0, 1) > |W_n^{obs}|) = 2(1 - \Phi(|W_n^{obs}|)),$$

donde Φ denota la función de distribución acumulada normal estándar.

Observación 5.2 (Velocidad de convergencia). La corrección de continuidad (reemplazar T^+ por $T^+ \pm \frac{1}{2}$ según la dirección del contraste) mejora la aproximación normal para n moderado, dado que T^+ es un estadígrafo entero.