Asymptotic Normality of the Wilcoxon Signed-Rank Statistic via Lindeberg CLT

chatgpt.com

April 14, 2026

1 Model and notation

Let $X_1, \dots, X_n$ be i.i.d. continuous random variables with a distribution symmetric about 0. Define

$$R_i = \mathrm{rank}(|X_i|) \in \{1, \dots, n\}, \quad S_i = \mathbf{1}(X_i > 0).$$

The Wilcoxon signed-rank statistic is

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

Under symmetry:

- $\mathbb{P}(S_i = 1) = \mathbb{P}(S_i = 0) = 1/2$,
- (S_i) are independent of $(|X_i|)$ and hence of (R_i),
- $(R_1, \dots, R_n)$ is a uniform random permutation of $\{1, \dots, n\}$.

2 Expectation and variance

We use linearity and exchangeability.

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

The variance is known to be

$$\mathrm{Var}(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

3 Representation as a linear rank statistic

Define centered signs:

$$Z_i = S_i - \frac{1}{2}, \quad \mathbb{E}[Z_i] = 0, \quad \mathrm{Var}(Z_i) = \frac{1}{4}.$$

Then

$$T^+ - \mathbb{E}[T^+] = \sum_{i=1}^{n} R_i Z_i.$$

Conditionally on $R_1, \ldots, R_n$, this is a sum of independent mean-zero random variables.

Define weights

$$a_{ni} = R_i.$$

Then

$$T^+ - \mathbb{E}[T^+] = \sum_{i=1}^{n} a_{ni} Z_i.$$

This is a triangular array structure suitable for the Lindeberg-Feller CLT.

4 Variance normalization

Let

$$V_n = \mathrm{Var}(T^+) = \frac{n(n+1)(2n+1)}{24} \sim \frac{n^3}{12}.$$

Define normalized sum

$$U_n = \frac{T^+ - \mathbb{E}[T^+]}{\sqrt{V_n}}.$$

Then

$$U_n = \sum_{i=1}^{n} b_{ni} Z_i, \quad b_{ni} = \frac{R_i}{\sqrt{V_n}}.$$

5 Lindeberg condition

We verify the Lindeberg condition for the triangular array $\{b_{ni} Z_i\}$.

Since $|Z_i| \leq 1/2$, we have for any $\varepsilon > 0$:

$$|b_{ni} Z_i| \leq \frac{R_i}{2\sqrt{V_n}} \leq \frac{n}{2\sqrt{V_n}} \to 0.$$

Hence for large n, no single term contributes a large amount, and Lindeberg's condition is trivially satisfied:

$$\sum_{i=1}^{n} \mathbb{E}[b_{ni}^2 Z_i^2 \mathbf{1}(|b_{ni} Z_i| > \varepsilon)] = 0 \quad \text{for large } n.$$

6 Asymptotic variance normalization

We compute:

$$\sum_{i=1}^{n} \mathbb{E}[b_{ni}^2 Z_i^2] = \frac{1}{4V_n} \sum_{i=1}^{n} \mathbb{E}[R_i^2].$$

Since R_i is uniform on $\{1, \ldots, n\}$:

$$\mathbb{E}[R_i^2] = \frac{(n+1)(2n+1)}{6}.$$

Thus

$$\sum_{i=1}^{n} \mathbb{E}[b_{ni}^2 Z_i^2] = \frac{n(n+1)(2n+1)}{24V_n} = 1.$$

7 Application of Lindeberg-Feller CLT

Since:

- Z_i are independent, mean zero,
- the variance normalization holds,
- Lindeberg condition is satisfied,

we conclude:

$$U_n \xrightarrow{d} \mathcal{N}(0, 1).$$

8 Conclusion

Therefore,

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

This establishes the asymptotic normality of the Wilcoxon signed-rank statistic.