Notes on Nonparametric Statistics

José Manuel Rodríguez Caballero

September 7, 2024

1 The Wilcoxon Rank Sum

Given two positive integers n and N, we define the Wilcoxon Rank Sum distribution $\mathbf{P}\left(W_n^N = k\right)$ via the generating function

$$\frac{\binom{N}{n}_q}{\binom{N}{n}} = \sum_{k=\frac{n(n+1)}{2}}^{\frac{n(2N-n+1)}{2}} \mathbf{P}\left(W_n^N = k\right) q^k, \tag{1}$$

where the polynomials $\binom{N}{n}_q$ are given by the product

$$\prod_{n=1}^{N} (1 - zq^n) = \sum_{n=0}^{N} (-1)^n \binom{N}{n}_q z^n. \tag{2}$$

Theorem 1. *For two positive integers n and N, we have*

$$\mathbf{E}\left(W_n^N\right) = \frac{n(N+1)}{2}, \tag{3}$$

$$\mathbf{Var}\left(W_n^N\right) = \frac{n(N-n)(N+1)}{12}. \tag{4}$$

Proof. Using the notations

$$F = \prod_{n=1}^{N} (1 - zq^n), \tag{5}$$

$$G = \sum_{k \geq 1} \left(q^k + q^{2k} + q^{3k} + \dots + q^{Nk}\right) \frac{z^k}{k}, \tag{6}$$

we have

$$F = e^{-G}. \tag{7}$$

Let $\mathcal{H} = q\frac{d}{dq}$, it follows

$$\begin{aligned} \mathcal{H}F &= -e^{-G}\mathcal{H}, & (8) \\ \mathcal{H}^2 F &= e^{-G}\left[(\mathcal{H}G)^2 - \mathcal{H}^2 G\right]. & (9) \end{aligned}$$

We compute

$$\begin{aligned} \mathcal{H}G\big|_{q=1} &= \frac{N(N+1)}{2}\frac{z}{1-z}, & (10) \\ \mathcal{H}^2 G\big|_{q=1} &= \frac{N(N+1)(2N+1)}{6}\frac{z}{(1-z)^2}, & (11) \end{aligned}$$

and

$$\begin{aligned} \mathcal{H}F\big|_{q=1} &= -\frac{N(N+1)}{2}z(1-z)^{N-1}, & (12) \\ \mathcal{H}^2 F\big|_{q=1} &= (1-z)^{N-2}\left[\left(\tfrac{N(N+1)}{2}\right)^2 z^2 - \tfrac{N(N+1)(2N+1)}{6}z\right]. & (13) \end{aligned}$$

Considering the coefficients of both sides,

$$\begin{aligned} \frac{\mathcal{H}\binom{N}{n}_q\Big|_{q=1}}{\binom{N}{n}} &= \frac{n(N+1)}{2}, & (14) \\ \frac{\mathcal{H}^2\binom{N}{n}_q\Big|_{q=1}}{\binom{N}{n}} &= \frac{n(N+1)(3nN+2n+N)}{12}. & (15) \end{aligned}$$

By definition of the expected value,

$$\begin{aligned} \frac{\mathcal{H}\binom{N}{n}_q\Big|_{q=1}}{\binom{N}{n}} &= \mathbf{E}\left[W_n^N\right], & (16) \\ \frac{\mathcal{H}^2\binom{N}{n}_q\Big|_{q=1}}{\binom{N}{n}} &= \mathbf{E}\left[\left(W_n^N\right)^2\right]. & (17) \end{aligned}$$

Combining (14) and (16), we obtain (3).

By definition of the variance,

$$\mathbf{Var}\left[W_n^N\right] = \mathbf{E}\left[\left(W_n^N\right)^2\right] - \left(\mathbf{E}\left[W_n^N\right]\right)^2. \tag{18}$$

Combining (14), (15), (16) and (18) we obtain (4). □

Theorem 2. *For two positive integers n and N, we have that W_n^N is symmetric, i.e.,*

$$\mathbf{P}\left(W_n^N = k\right) = \mathbf{P}\left(W_n^N = n\left(N+1\right) - k\right), \tag{19}$$

for all $\frac{n(n+1)}{2} \leq k \leq \frac{n(2N-n+1)}{2}$.

Proof. The function $F(q,z) = \prod_{n=1}^{N}\left(1 - zq^n\right)$ is invariant under the transformation $q \mapsto q^{-1}$ and $z \mapsto zq^{N+1}$,

$$F\left(q^{-1}, zq^{N+1}\right) = F\left(q, z\right). \tag{20}$$

Equating coefficients in (20), according to (2), we obtain

$$q^{n(N+1)}\binom{N}{n}_{q^{-1}} = \binom{N}{n}_q, \tag{21}$$

where $\binom{N}{n}_{q^{-1}}$ is the result of the substitution of q by q^{-1} in $\binom{N}{n}_q$. Combining (21) and (1), we conclude (19). □

Theorem 3. *For two positive integers n and N, consider two families of random variables $X_1, X_2, X_3, ..., X_{N-n}$ and $Y_1, Y_2, Y_3, ..., Y_n$ such that they together are a random permutation of $1, 2, 3, ..., N$. For each $0 \leq k \leq n(N-n)$, we have*

$$\mathbf{P}\left(W_n^N = \frac{n(n+1)}{2} + k\right) = \mathbf{P}\left(\#\{(i,j) : X_i < Y_j\} = k\right). \tag{22}$$

Proof. Without lost of generality, assume that $X_1 < X_2 < X_3 < ... < X_{N-n}$ and $Y_1 < Y_2 < Y_3 < ... < Y_n$. Let k be the number of (i,j) such that $X_i < Y_j$. Notice that $0 \leq k \leq n(N-n)$, because there are $N-n$ values of X_i and n values of Y_j. Our goal is to compute

$$W_n^N = Y_1 + Y_2 + Y_3 + ... + Y_n. \tag{23}$$

For each $1 \leq j \leq n$, the number of i such that $X_i < Y_j$ is $(Y_j - 1) - (j - 1) = Y_j - j$. Hence,

$$\sum_{j=1}^{n} (Y_j - j) = k. \tag{24}$$

It follows that

$$W_n^N = \frac{n(n+1)}{2} + k. \tag{25}$$

Therefore, (22) holds. $\square$