axioms

Article

Smooth Versions of the Mann–Whitney–Wilcoxon Statistics

Netti Herawati [1,*] and Ibrahim A. Ahmad [2]

[1] Department of Mathematics, University of Lampung, Lampung 35141, Indonesia
[2] Department of Statistics, Oklahoma State University, Stillwater, OK 74078, USA; drabe_aa@yahoo.com
* Correspondence: netti.herawati@fmipa.unila.ac.id

Abstract: The well-known Mann–Whitney–Wilcoxon (MWW) statistic is based on empirical distribution estimates. However, the data are often drawn from smooth populations. Therefore, the smoothness characteristic is not preserved. In addition, several authors have pointed out that empirical distribution is often an inadmissible estimate. Thus, in this work, we develop smooth versions of the MWW statistic based on smooth distribution function estimates. This approach preserves the data characteristics and allows the efficiency of the procedure to improve. In addition, our procedure is shown to be robust against a large class of dependent observations. Hence, by choosing a rectangular array of known distribution functions, our procedure allows the test to be a lot more reflective of the real data.

Keywords: Mann–Whitney–Wilcoxon; strong and weak consistency; asymptotic normality; robustness

Citation: Herawati, N.; Ahmad, I.A. Smooth Versions of the Mann–Whitney–Wilcoxon Statistics. *Axioms* **2022**, *11*, 300. https://doi.org/10.3390/axioms11070300

Academic Editor: Hans J. Haubold

Received: 27 May 2022
Accepted: 15 June 2022
Published: 21 June 2022

1. Introduction

Suppose that X and Y are two independent random variables with the distribution functions (df's) F and G, respectively. We can say that stochastically $X \leq Y$ if $F(x) \geq G(x)$ for all x. Testing $H_0: F(x) = G(x)$ for all x against the alternative $H_1: F(x) \leq G(x)$ is carried out via the celebrated Mann–Whitney–Wilcoxon (MWW) statistics. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two samples from F and G, respectively. The MWW statistics are defined by the following:

$$p_{m,n} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{\{X_i < Y_j\}} \quad (1)$$

where $I_A(a) = 1$ if $a \in A$ and 0 otherwise [1,2]. Note that $p_{m,n}$ is the empirical estimate of $p = P(X \leq Y) = \int F(x) dG(x)$. Several generalizations of $p_{m,n}$ were discussed in the literature, with the aim of increasing the test efficiency, c.f. [3–5].

On the other hand, ref. [6] showed that $F_m(x)$ with respect to the integrated mean squared error (IMSE) is inadmissible. Therefore, extensive research has been carried out to find competing estimates of $F(x)$ and $F_m(x)$. Additionally, work on estimating the corresponding probability density function (pdf) began with the pioneering work of [7,8], who suggested that a smooth estimate of $F(x)$ can be gained by integrating the so-called "kernel pdf estimates" defined by the following:

$$\hat{f}(x) = \frac{1}{m a_m} \sum_{i=1}^{m} k\left(\frac{x - X_i}{a_m}\right), \quad (2)$$

where $k(\cdot)$ is a known pdf with the sequence of real numbers of $\{a_m\}$, such that $a_m \to 0$ as $m \to \infty$. Hence, the kernel of estimate of $F(x)$ is as follows:

$$\hat{F}_m(x) = m^{-1} \sum_{i=1}^{m} K\left(\frac{x - X_i}{a_m}\right), \tag{3}$$

where $K(u) = \int_{-\infty}^{\infty} k(w)dw$.

A large number of studies have been carried out to investigate the properties of $\hat{F}_m(x)$, including the works of [9–13], among many other authors.

The idea of comparing the efficiency of $\hat{F}_m(x)$ to that of $F_m(x)$ started with the work of [14], who showed that the relative deficiency of $F_m(x)$ to an appropriately selected $\hat{F}_m(x)$ quickly tended to ∞ as $m \to \infty$ when using the MISE as a criterion. This was followed by a large number of authors, among whom we mention [15–17], who showed the following:

$$\int_{-\infty}^{\infty} E\{\hat{F}_m(x) - F(x)\}^2 \, dF(x) = \frac{1}{6m} - 2a_m m^{-1} C \int_{-\infty}^{\infty} f^2(x)\, dx + \frac{h^4 \sigma_k^2}{4} \int_{-\infty}^{\infty} \left(f'(x)\right)^2 f(x) dx + o(a_m/m + a_m^4), \tag{4}$$

where $C = \int_{-\infty}^{\infty} t\, k(t)\, K(t) dt$ and σ_k^2 is the variance of $k(\cdot)$. For further discussion, see the works by [18–25], among others. In cases where the data are censored, we refer to the works of [26,27].

A natural extension of Equation (3) can be defined as follows:

Let $\{W_m\}$ be a sequence of known df's. We define an extended estimate of $F(x)$ using the following:

$$\hat{F}_m(x) = m^{-1} \sum_{i=1}^{m} W_m\,(x - X_i), \tag{5}$$

where $W_m(u) \to I_{(-\infty,\infty)}(u)$ as $m \to \infty$. Note that the kernel df estimate takes $W_m(u) = K(\frac{u}{a_m})$.

Our smooth estimate of p is defined as follows: let $\{W_{m,n}\}$ be a rectangular array of known df's satisfying the following:

$$W_{m,n}(u) \to I_{(-\infty,\infty)}(u) \text{ as } \min(m.n) \to \infty \tag{6}$$

We propose estimating p using the following:

$$\hat{p}_{m,n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{m,n}\,(y - x) dF_m(x) dG_n(y) = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} W_{m,n}\,(Y_j - X_i) \tag{7}$$

Note that we can write $\hat{p}_{m,n}$ as $\hat{p}_{m,n} = n^{-1} \sum_{j=1}^{n} \hat{F}_m\left(Y_j\right)$, where $\hat{F}_m(x) = m^{-1}$ and $\sum_{i=1}^{m} W_{m,n}\,(x - X_i)$. Examples of the $W_{m,n}(u)$ arrays include the following:

1. $W_{m,n}(u) = \frac{1}{2}\,[W_m(u) + W_n(u)]$;
2. $W_{m,n}(u) = \int_{-\infty}^{\infty} W_m(u - v) dW_n(v)$;
3. $W_{m,n}(u) = [W_m(u) W_n(u)]^{1/2}$.

2. Results

2.1. Large Sample Theory of $\hat{p}_{m,n}$

(i) Asymptotic unbiasedness: if $W_{m,n}(u) \to I_{(0,\infty)}(u)$ as $\min(m, n) \to \infty$, then $\mathrm{E}\hat{p}_{m,n} \to p$.

Proof. Note that $\mathrm{E}\hat{p}_{m,n} = \iint W_{mn}(y - x) dF(x) dG(y)$ and let $H_{m,n}(y) = \int W_{m,n}(y - x) dF(x)$. Thus, the characteristic function of $H_{m,n}(y)$ is as follows:

$$\emptyset_{H_{m,n}}(t) = \emptyset_{W_{m,n}}(t) \emptyset_F(t), \tag{8}$$

where $\emptyset_{W_{m,n}}(\emptyset_F)$ denotes the characteristic function of $W_{m,n}\,(F)$. Since $W_{m,n}(t) \to I_{(0,\infty)}(x)$, $\emptyset_{W_{m,n}}(t) \to 1$ as $\min(\mathrm{m,n}) \to \infty$ and, thus, $\emptyset_{H_{m,n}}(t) \to \emptyset_F(t)$. Hence, $H_{m,n}(\cdot) \to F(\cdot)$ as $\min(m, n) \to \infty$ at each continuity point of F. The Lebesgue dominated convergence theorem was applied to obtain the result. □

(ii) Weak (and L_1) consistency: if $W_{m,n}(u) \to I_{(0,\infty)}(u)$ as $\min(\mathrm{m,n}) \to \infty$, then $\hat{p}_{m,n} \to p$ with the same probability as $\min(m,n) \to \infty$.

Proof. From Part (i), we only need to show that $V(\hat{p}_{m,n}) \to 0$ as $\min(m,n) \to \infty$. We will show a stronger result for sufficiently large values of m and n, as follows:

$$V(\hat{p}_{m,n}) \simeq \frac{2}{m}\iint_{y_i<y_2} F(y_1)\,[1-F(y_2]dG(y_1)dG(y_2) + \frac{2}{n}\iint_{x_i<x_2} G(x_1)[1-G(x_2]dF(x_1)dF(x_2) \;= \sigma^2_{m,n} \tag{9}$$

Note that

$$\begin{aligned} V(\hat{p}_{m,n}) &= (mn)^{-1}EW^2_{m,n}(Y_1-X_1) + [(m-1)/mn]EW_{m,n}(Y_1-X_1)W_{m,n}(Y_1-X_2) \\ &+[(n-1)/mn]EW_{m,n}(Y_1-X_1)W_{m,n}(Y_2-X_1) - [(m+n-1)/mn]E\big[W_{m,n}(Y_1-X_1)\big]^2 \\ &= A_1+A_2+A_3+A_4 \end{aligned} \tag{10}$$

Since $W^2_{m,n}(u) \to I_{(0,\infty)}(u)$ as $\min(m,n) \to \infty$, using an argument similar to Part (i), we can show that $\int W^2_{m,n}(y-x)dF(x) \to F(y)$ as $\min(m,n) \to \infty$ at each continuity point y of F_1. Thus, it follows that $A_1 \to 0$ as $\min(m,n) \to \infty$. In fact, $A_1 = o[(mn)^{-1}]$. Next, for sufficiently large values of m and n, we can see the following by the same reasoning:

$$\begin{aligned} A_2 &\simeq [(m-1)/mn]\iiint W_{m,n}(y-x_1)\,W_{m,n}(y-x_2)dF(x_1)dF(x_2)dG(y) \\ &= n^{-1}\iint G[\max\,(x_1,x_2)]dF(x_1)dF(x_2), \end{aligned} \tag{11}$$

For sufficiently large values of m and n, $\int W_{m,n}(y-x_1)W_{m,n}(y-x_2)\,d\,G(y) \simeq G[\max(x_1,x_2)]$. The proof for A_3 is similar. Finally, for $A_4 \simeq [(m+n-1)/mn]p^2$, the desired conclusion is reached by collecting terms. □

(iii) Strong consistency: if $W_{m,n}(u) \to I_{(0,\infty)}(u)$ as $\min(\mathrm{m,n}) \to \infty$, then $\hat{p}_{m,n} \to p$ with a probability of one.

Proof. Since, according to Part (i), $\mathrm{E}\hat{p}_{m,n} \to p$ as $\min(\mathrm{m,n}) \to \infty$, we only need to look at $[\hat{p}_{m,n} - E\hat{p}_{m,n}]$. However, since $W_{m,n}(\cdot)$ is a distribution function, by integrating the parts, we obtain the following:

$$\begin{aligned} |\hat{p}_{m,n} - E\hat{p}_{m,n}| &= \left|\iint W_{m,n}(y-x)dF_m(x)dG_n(y) - \iint W_{m,n}(y-x)dF(x)dG(y)\right| \\ &\le \left|\iint W_{m,n}(y-x)dF_m(x)dG_n(y) - \iint W_{m,n}(y-x)dF_m(x)dG(y)\right| \\ &+ \left|\iint W_{m,n}(y-x)dF_m(x)dG(y) - \iint W_{m,n}(y-x)dF(x)dG(y)\right| \\ &= \left|\iint [F_m(x)-F(x)]dW_{m,n}(y-x)dG_n(y)\right| + \left|\iint [G_n(y)-G(y)]dW_{m,n}(y-x)dF(x)\right| \\ &\le \sup_x|F_m(x)-F(x)| + \sup_y\,|G_n(y)-G(y)| \\ &= o\left(m^{-\frac{1}{2}}\sqrt[2]{ln\,ln\,m}\right) + o\left(n^{-\frac{1}{2}}\sqrt[2]{ln\,ln\,n}\right) \\ &= o\left\{\min\sqrt{\frac{\ln\ln m}{m}},\sqrt{\frac{\ln\ln n}{n}}\right\} = o(1) \end{aligned} \tag{12}$$

We use the standard law of iterated logarithm for empirical distribution functions. □

(iv) Asymptotic normality: if $W_{m,n}(u) \to I\ \ (u)$, and if $\sqrt{\min(m,n)}(E\hat{p}_{mn} - p) \to 0$ as $\min(\mathrm{m,n}) \to \infty$, $(\hat{p}_{m,n} - p)/\sigma_{m,n}$ is asymptotically standard normal, where $\sigma_{m,n}$ is as given in Equation (9).

Proof. Under the above conditions, it is sufficient to point out that $(\hat{p}_{m,n} - E\hat{p}_{m,n})/\sigma_{m,n}$ is asymptotically standard normal. To this end, we obtain the following:

$$\begin{aligned}\hat{p}_{m,n} - E\hat{p}_{m,n} &= \left\{\iint [G_n(y) - G(y)]dW_{m,n}(y-x)dF(x)\right.\\ &\quad \left.- \iint [F_m(x) - F(x)]dW_{m,n}(y-x)dG(y)\right\}\\ &\quad + \left\{\iint [F_m(x) - F(x)]dW_{m,n}(y-x)d[G_n(y) - G(y)]\right\}\\ &= B + C\end{aligned} \tag{13}$$

Clearly, B is the difference between the two independent sample averages; thus, we have proved that B is asymptotically normal and has a mean of zero with the variance $\sigma^2_{m,n}$. Additionally, note that $EC = o(1)$, and if m and n are large enough according to the methods in Part (i), then we obtain the following:

$$EC^2 \simeq (mn)^{-1}\left\{[\int F(x)(1-F(x)dG(x)] - 2\iint_{x_i<x_2} F(x_1)(1-F(x_2)dh(x_1)dh(x_2)\right\} = o[(mn)^{-1}].$$

Hence, $\min(m,n)\,EC^2 = o[(\min(m,n))^{-1}] = o(1)$, thus $\sqrt{\min(m,n)}\,C \to 0$. The conclusion is now obtained. □

Some remarks:

(I) A sufficient condition for $\sqrt{\min(m,n)}(E\hat{p}_{m,n} - p) \to 0$ is that $\sqrt{\min(m,n)}\int |t|^\alpha\, dW_{m,n}(t) \to 0$ as $\min(m,n) \to \infty$, provided that F is the Lipschitz of order $0 < \alpha \le 1$.

Proof. Note that $E\hat{p}_{m,n} = \int E\hat{F}_{m,n}(y)dG(y)$, where $E\hat{F}_{m,n}(y) = \int W_{m,n}(y-t)dF(t)$. Thus, by integrating the parts, the following are obtained:

$$\begin{aligned}&\left|\iint W_{m,n}(y-x)dF(x)dG(y) - \int F(x)dG(x)\right|\\ &= \left|\int E\hat{F}_{m,n}(x)dG(x) - \int F(x)dG(x)\right|\\ &\le \iint \left|F(x-t) - F(x)dW_{m,n}(t)dG(x)\right|\\ &\le C\int (t)^\alpha\, dW_{m,n}(t) = o\left[\left(\sqrt{\min(m,n)}\right)^{-1}\right].\ \square\end{aligned}$$

If we know that $W_{m,n}(t) = \left[K\left(\frac{t}{a_m}\right) + K\left(\frac{t}{b_n}\right)\right]/2$ with a known $K(\cdot)$ and $k(t) = \frac{d}{dt}K(t)$, $a_m = C_1 m^{-\delta}$, $C_1 > 0$ and $\frac{1}{2} < \delta < 1$, and $b_n = C_2 n^{-\beta}$, $C_2 > 0$ and $\frac{1}{2} < \beta < 1$, then the above condition is met for $\alpha = 1$ and $\delta = \beta$. Of course, if $W_{m,n}(u) = I_{(0,\infty)}(u)$, then $\sqrt{\min(m,n)}\,(E\hat{p}_{mn} - p) = 0$.

(II) If one wishes to find an asymptotic confidence interval for $p = P(X < Y)$, a consistent estimate of $\sigma^2_{m,n}$ is needed. However, this estimate can be easily obtained using the following:

$$\begin{aligned}\hat{\sigma}^2_{m,n} &= \frac{2}{m}\iint_{y_1<y_2} \hat{F}_m(y_1)\left(1 - \hat{F}_m(y_2)\right)dG_n(y_1)dG_n(y_2)\\ &\quad + \frac{2}{n}\iint_{x_1<x_2} \hat{G}_n(x_1)(1 - \hat{G}_n(x_2))dF_m(x_1)dF_m(x_2)\end{aligned} \tag{14}$$

where $\hat{F}_m(x) = m^{-1}\sum_{i=1}^m W_{m,n}(x - X_i)$ and $F_m(x)$ is the empirical df of F, with $\hat{G}_n(y)$ and $G_n(y)$ defined analogously. Thus, we obtain the confidence bounds as follows:

$$\hat{p}_{m,n} \pm Z_{\alpha/2}\,\hat{\sigma}_{m,n}/[\min(m,n)]^{1/2} \tag{15}$$

In addition, note that $\sigma^2_{m,n} \simeq [\min(m,n)]^{-1}\left\{\iint_{y_1<y_2} F(y_1)\,(1 - F(y_2))dG(y_1)dG(y_2) + \iint_{x_1<x_2} G(x_1)\,(1 - G(x_2))dF(x_1)dF(x_2)\right\} = [\min(m,n)]^{-1}\sigma^2$. Thus, in Part (iii) above, we can write that $(\hat{p}_{m,n} - p)\,[\min(m,n)]^{1/2}/\hat{\sigma}^2_{m,n}$ is asymptotically standard normal.

(III) The kernel method provides an easy way to generate $W_{m,n}(\cdot)$, at least for the special cases $W_{m,n}(u) = \frac{1}{2}\,[W_m(u) + W_n(u)]$ and $W_{m,n}(u) = \int_{-\infty}^{\infty} W_m(u-v)\, dW_n(v)$, by defining $W_m(W_n)$ as a kernel estimate V_{ij}., $W_m(u) = K\left(\frac{u}{a_m}\right), W_n(u) = K\left(\frac{u}{b_n}\right)$, where K is a random df and $\{a_m\}(\{b_n\})$ is a sequence of constants, such that $a_m \to 0$ as $m \to \infty$ ($b_n \to 0$ as $n \to \infty$). As is now well known in the literature, the choices of a_m and b_n will, generally speaking, depend on the data. Thus, in such cases, the conditions on $W_{m,n}(u)$ will have to be adjusted.

(IV) Mann–Whitney–Wilcoxon statistics for paired data. $(X_1, Y_1), \ldots, (X_n, Y_n)$ is drawn from a bivariate df $F(x, y)$. In this case, we may be interested in estimating either (i) $P(X_1 < Y_1)$ or (ii) $P(X_1 < Y_2)$. We propose estimating $p_1 = P(X_1 < Y_1)$ using $\hat{p}_1 = \iint W_n(x-y) dF_n(x,y) = n^{-1} \sum_{i=1}^{n} W_n(X_i - Y_i)$, where $\{W_n\}$ is a sequence of df's converging to $I_{(0,\infty)}(\cdot)$ as $n \to \infty$. Note that $\hat{p}_1$ is an average of independent random variables. Thus, one can, without much difficulty, study its properties. We have left in the details for readers who are interested. If it is necessary to estimate $p_2 = P(X_1 < Y_2)$, we propose estimating $\hat{p}_2 = n(n-1) \sum_{i \neq j} W_n(Y_j - X_i)$. Thus, $\hat{p}_2 = \hat{p}_n - \hat{p}_1$, where $\hat{p}_n = \iint W_n(x - y) dF_n(x) dG_n(y) = n^{-2} \sum_i \sum_j W_n(Y_j - X_i)$. The asymptotic properties of $\hat{p}_2$ are not trivial to deduce, but can be obtained by the methods described in this paper. The consistency (weak and strong) of $\hat{p}_2$ when we have a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is obtained under the condition that $W_n(\cdot) \to I_{(o,\infty)}(\cdot)$ as $n \to \infty$. The asymptotic normality is obtained with the following approximate variance:

$$\sigma^2 = \int G^2(x)\, dF(x) + \int \left(1 - F(x)\right)^2 dG(x) + 2 \iint G(x)\left(1 - F(x)\right) dH(x,y) - 4\hat{p}_2^2 \tag{16}$$

To estimate $\sigma_{m,n}^2$, we propose the following:

$$\hat{\sigma}_{m,n}^2 = \int G_n^2(x)\, dF_m(x) + \int \left(1 - F_m(x)\right)^2 dG_n(x) + 2 \iint G_n(x)\left(1 - F_m(x)\right) dH_n(x,y) - 4\hat{p}_2^2$$

Thus, $\sqrt{n}(\hat{p}_2 - p_2)/\hat{\sigma}_{m,n}$ is asymptotically normal provided that $\sqrt{n}\,(E\,\hat{p}_2 - p_2) \to 0$ as $n \to \infty$.

2.2. Robustness of $\hat{p}_{mn}$ against Dependence

In this section, we assume that $X_1, \ldots, X_m (Y_1, \ldots, Y_n)$ denotes the first m(n) units in the sequence $\{X_m\}(\{Y_n\})$, satisfying the following strong mixing condition c.f. [28]. Let $\Im_a^b$ denote the σ-field generated by $X_a, \ldots, X_b$, then $\{X_m\}$ is said to be strong mixing if there is a function with the integer value $\alpha(\cdot)$, such that $\alpha(m) \to 0$ as $m \to \infty$, and the following is obtained:

$$|P(AB) - p(A)p(B)| \leq \alpha(m) \tag{17}$$

For all $A \in \Im_{-\infty}^{a}, B \in \Im_{a+m}^{\infty}$. Throughout this section, we shall assume that $\{X_m\}(\{Y_n\})$ are strictly stationary. To establish the results of this section, we need some definitions.

Let $U_{m,n}(X_1) = E[W_{m,n}(Y_1 - X_1)|\, X_1]$ and $V_{m,n}(Y_1) = E[W_{m,n}(Y_1 - X_1)|\, Y_1]$. The following are obtained:

$$\begin{gathered} R_{m,n} = [\min(m,n)]^{1/2} \left\{ m^{-1} \sum_{i=1}^{m} U_{m,n}(X_i) + n^{-1} \sum_{j=1}^{n} V_{m,n}(Y_j) \right. \\ \left. -(mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} W_{m,n}(Y_j - X_i) - EW_{m,n}(Y_j - X_i) \right\} \end{gathered} \tag{18}$$

The next lemma is instrumental in the development of this section.

Lemma 1. *If* $\sum_{m=1}^{\infty} \alpha(m) < \infty$ *and* $\sum_{n=1}^{\infty} \alpha(n) < \infty$, *then* $ER_{m,n} \to 0$ *as* $min(m,n) \to \infty$.

Proof. For $1 \leq i \leq m$ and $1 \leq j \leq n$, the following is obtained:

$$\varphi_{m,n}(i,j) = U_{m,n}(X_i) + V_{m,n}(Y_j) - W_{m,n}(Y_j - X_i) - EW_{m,n}(Y_j - X_i). \tag{19}$$

Then, we can easily see the following:

$$ER_{m,n}^2 = \frac{\min(m,n)}{(mn)^2} E\left\{\sum_i \sum_j \varphi_{m,n}(i,j)\right\}^2$$
$$= \frac{\min(m,n)}{(mn)^2}\left\{\sum_i \sum_j E\varphi_{m,n}^2(i,j) + \sum_i \sum_{j\neq j^*} E\varphi_{m,n}(i,j)\varphi_{m,n}(i,j^*)\right\}$$
$$+ \sum_{i\neq i^*} \sum_j E\,\varphi_{m,n}(i,j)\varphi_{m,n}(i^*,j) + \sum_{i\neq i^*} \sum_{j\neq j^*} E\,\varphi_{m,n}(i,j)\varphi_{m,n}(i^*,j^*). \tag{20}$$

We shall consider each term alone. Since $|\varphi(i,j)| \leq 2$, the first sum of the order $[\min(m,n)/mn] \to 0$ as $\min(m,n) \to \infty$. From now on, we drop the m and n suffix from φ. Next, by the Lemma of [28], since $\{X_i\}$ and $\{Y_j\}$ are strictly stationary, we can see the following:

$$\left|\sum_i \sum_{j\neq j^*} E\varphi(i,j)\varphi(i,j^*)\right| \leq 2\sum_i \sum_{j\neq j^*} |E\varphi(i,j)\varphi(i,j^*)|$$
$$= 2\sum_i \sum_j (n-j+1)|E\varphi(i,j)\varphi(i,j^*)|$$
$$\leq C\,m\sum_{j=1}^n (n-j+1)\,\alpha(j) = C\,mn\sum_{j=1}^n \left(\frac{n-j+1}{n}\right)\alpha(j)$$
$$\leq C\,mn\left[\sum_{j=1}^\infty \alpha(j) - \frac{1}{n}\sum_{j=1}^n j\,\alpha(j)\right] \tag{21}$$

Thus, the second sum in Equation (20) is bounded above by the following:

$$C\left[\frac{\min(m,n)}{mn}\right]\left[\sum_{j=1}^\infty \alpha(j) - \frac{1}{n}\sum_{j=1}^n j\,\alpha(j)\right], \tag{22}$$

which converges to zero, since $\sum_j \alpha(j) < \infty$. Thus, by Kronecker's lemma, $n^{-1}\sum_j j\,\alpha(j) \to 0$ as $n \to \infty$. In a similar way, we can show that the third term in Equation (20) converges to zero as $\min(m,n) \to \infty$. Finally, the fourth term is less than or equal to

$$4\sum_{i<i^*}\sum_{j<j^*} |E\varphi(i,j)\varphi(i^*,j^*)| = 4\sum_{i=1}^m\sum_{j=1}^n (m-i+1)(n-j+1)|E\varphi(1,1)\varphi(i+1,j+1)|.$$

However, since $|E\varphi(1,1)\varphi(i+1,j+1)| \leq C\,\alpha(i)$ and $|E\varphi(1,1)\varphi(i+1,j+1)| \leq C\,\alpha(j)$ (note that here and elsewhere, C denotes a generic positive constant that is not necessarily the same from place to place), $|E\varphi(1,1)\varphi(i+1,j+1)| \leq C\,\alpha(\max(i,j))$. Thus, the last term in Equation (20) is less than or equal to

$$C\sum_{i=1}^m\sum_{j=1}^n (m-i+1)(n-j+1)\,\alpha(\max(i.j)) \leq C\,mn\sum_i\sum_j \alpha(\max(i.j))$$
$$= C\,mn\left\{\sum_i\sum_{j=1}^i \alpha(i) + \sum_{i=1}^m\sum_{j=i+1}^n \alpha(j)\right\} \leq C\,mn\left\{\sum_{i=1}^m i\,\alpha(i) + \sum_{i=1}^m\left[\sum_{j=i+1}^\infty \alpha(j)\right]\right\}$$
$$= o(m^2 n) \tag{23}$$

To show Equation (23), let k_m be an n integer, such that $k_m < m$ and $k_m \to \infty$ as $m \to \infty$ and $k_m = o(m)$. Then, the following is obtained:

$$\sum_{i=1}^m i\,\alpha(i) = \sum_{i=1}^{k_m} i\,\alpha(i) + \sum_{i=k_m+1}^m i\,\alpha(i) \leq k_m\sum_1^\infty \alpha(i) + m\left[\sum_{i=k_m}^\infty \alpha(i)\right]$$
$$= o(k_m) + m\,o(1) = o(m). \tag{24}$$

Next, since $\sum_1^\infty \alpha(m) < \infty$, $\sum_{i=1}^m\left[\sum_{j=i+1}^\infty \alpha(j)\right] = o(m)$ as $m \to \infty$. Thus, Equation (23) is proved, as is the lemma. □

In light of Lemma 1, the consistency and asymptotic normality of $\hat{p}_{m,n}$ can be analyzed just by looking at $m^{-1}\sum_{i=1}^m U_{m,n}(X_i) - E\,W_{m,n}(Y_1 - X_1)$ and $n^{-1}\sum_{j=1}^n V_{m,n}(Y_j) - EW_{m,n}(Y_1 - X_1)$. In addition, note that $EU_{m,n}(X_1) = EV_{m,n}(Y_1) = EW_{m,n}(Y_1 - X_1)$. Now, by stationarity, the following is obtained:

$$V\left[m^{-1}\sum_{i=1}^{m}U_{m,n}(X_i)\right] = m^{-1}\left\{\sum_{i=1}^{m}V\left(U_{m,n}(X_i)\right) + \sum_{i\neq i^*}cov(U_{m,n}(X_i),U_{m,n}(X_{j^*}))\right\}$$

$$= m^{-1}\,V\left(U_{m,n}(X_i)\right) + \frac{2}{m^2}\sum_{i=1}^{m}(m-i+1)\,cov(U_{m,n}(X_i),U_{m,n}(X_{i+1}))$$

Similarly, we can express $V(n^{-1}(\sum_{j=1}^{n}V_{m.n}(Y_j))$. Thus, we obtain the following for sufficiently large values of m and n:

$$V\left[m^{-1}\sum_{i=1}^{m}U_{m,n}(X_i) + n^{-1}\sum_{j=1}^{n}V_{m,n}\ (Y_j)\right]$$
$$\simeq \sigma_{m,n}^2 + \frac{2}{m^2}\sum_{i=1}^{m}(m-i+1)\,cov(U_{m,n}(X_1),U_{m,n}(X_{i+1})) \quad (25)$$
$$+\frac{2}{n^2}\sum_{j=1}^{n}(n-j+1)\,cov(V_{m,n}(Y_1),V_{m,n}(Y_{j+1)})$$

Let us find large sample values for the covariances. For large values of m and n, the following is obtained:

$$cov(U_{m,n}(X_1),U_{m,n}(X_{i+1)}) = \int U_{m,n}(X_1),U_{m,n}(X_{i+1)})dF_i(x_1,x_{i+1}) - \left[EW_{m,n}(Y_1 - X_1)\right]^2$$

$$\simeq \iint [F_i(x_1,x_{i+1}) - F(x_1)F(x_{i+1})]dG(x_i)dG((x_{i+1}) = \sigma_{1i}\,(x),$$

where $F_i(\cdot,\cdot)$ denotes the joint df of (X_1,X_{i+1}). This is similar for the second covariance term. Hence, for m and n values that are sufficiently large, the following is obtained:

$$V(\hat{p}_{mn}) \simeq \sigma_{m,n}^2 + \frac{2}{m^2}\sum_{i=1}^{m}(m-i+1)\,\sigma_{1i}(X) + \frac{2}{n^2}\sum_{j=1}^{n}(n-j+1)\,\sigma_{1j}(Y)$$
$$= \textstyle\sum_{m,n}^{2},\ \text{say.} \quad (26)$$

Now, let us write $\sum_{m,n}^2 = \sum_{m,n}^2(X) + \sum_{m,n}^2(Y)$, where $\sum_{m,n}^2(X)\left[\sum_{m,n}^2(Y)\right]$ is the approximate variance of $m^{-1}\sum_{i=1}^{m}U_{m,n}(X_i)\left[n^{-1}\sum_{j=1}^{n}V_{m,n}(Y_j)\right]$. Note that $\left|U_{m,n}(X_i)\right| \le 1$ and $\left|V_{m,n}(Y_j)\right| \le 1$. Thus, by Lemma 2.1 of [29], the following is obtained:

$$P\left[\left|m^{-1}\sum_{i=1}^{m}U_{m,n}(X_i) - EU_{m,n}(X_i)\right| \ge t\right] \le 2(1+K\,\alpha(M))^p e^{-Pt^2/2} \quad (27)$$

where $P = p(m)$ and $M = M(m)$ are integer-valued functions, such that $2PM \le m$.

Now we can state and prove the properties of $\hat{p}_{mn}$ if samples are drawn from strictly stationary strong mixing processes.

(i) Weak consistency: if $\sum_m \alpha(m) < \infty$ and $\sum_n \alpha(n) < \infty$, and if $W_{m,n}(u) \to I_{(0,\infty)}(u)$ as $\min(\text{m},\text{n}) \to \infty$, then $\hat{p}_{m,n} \to p$ with the same probability as $\min(\text{m},\text{n}) \to \infty$.

Proof. This directly follows the fact that $\hat{p}_{m,n} - E\,\hat{p}_{m,n} = \left(\bar{U}_{m,n}(X) - E\,\hat{p}_{m,n}\right) + \left(\bar{V}_{m,n}(Y) - E\hat{p}_{m,n}\right) + R_{m,n}/\min(m,n)$, and by applying the law of large numbers for Ergodic sequences, $\left(R_{m,n}/\min(m,n)\right) \to 0$ as $\min(m,n) \to \infty$, according to Lemma 1. □

(ii) Strong consistency: Assume that the conditions of Part (i) hold and, in addition, assume that there are integer-valued functions M(m) and $P(m)(N(n)$ and $Q(n))$, such that $2MP \le m(2NQ \le n)$ and for any $t > 0, \sum_{m=1}^{\infty}[1 + C\,\alpha(M)]^P e^{-Pt^2/2} < \infty[\sum_{n=1}^{\infty}(1 + C\,\alpha(N)]^Q e^{-Qt^2/2} < \infty$, then $\hat{p}_{m,n} \to p$ with a probability of one as $\min(m,n) \to \infty$.

Proof. Again, from Lemma 1, $E\left[\frac{R_{m,n}^2}{\min(m,n)}\right] = o[(min/m,n)^{-1}]^{-\gamma}$ for some $1 > \gamma > 0$ (by choosing based on the proof of that lemma, $R_m = m^\gamma$, $1 > \gamma > 0$). Thus, with $\upsilon = \min(m,n)$ and by writing R_υ for $R_{m,n}$, we can see that for any $\varepsilon > 0$, $\sum P[|R_\upsilon| > \varepsilon] < \infty$; thus, $R_\upsilon \to 0$ with a probability of one as $\upsilon \to \infty$. The conclusion is obtained by applying Equation (27) to $U_{m,n}(X_i)'s$ and $V_{m,n}(Y_j)'s$. □

(iii) Asymptotic normality: note that we obtain the following for large values of m, n:

$$Var(\hat{p}_{m,n}) \simeq [\min(m,n)]^{-1}\{\sigma^2 + \gamma(X) + \delta(Y)\} = [\min(m,n)]^{-1}\sigma^{*2}$$

where $\gamma(X) = \lim_{\min(m,n)\to\infty} \frac{2}{m}\sum_{i=1}^{m}(m-i+1)\sigma_{1i}(X)$. Similarly, we define $\delta(Y)$, provided, of course, that limits exist. In this case, we can write that $\sqrt{min(m,n)}\ (\hat{p}_{m,n} - p)/\sigma^*$ is asymptotically standard normal.

3. Discussion

This study has proven that the smooth version of Mann–Whitney–Wilcoxon statistics is robust against a large class of dependent observations, and can be used if the data are drawn from a smooth population. These procedures, which are based on a smooth distribution function, can maintain the nature of the data and allow the efficiency of the procedure to be increased. In addition, when selecting a rectangular array of known distribution functions, the smooth version of MWW statistics allows much better testing to be conducted. For future work, we shall apply MWW statistics to simulation data and real data.

Author Contributions: Conceptualization, N.H. and I.A.A.; methodology, N.H.; validation, I.A.A.; writing—original draft preparation, N.H. and I.A.A.; writing—review and editing, I.A.A.; supervision, I.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mann, H.B.; Whitney, D.R. On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60.
2. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1945**, *1*, 80–83.
3. Ahmad, I.A. A class of Mann-Whitney-Wilcoxon type statistics. *Am. Stat.* **1996**, *50*, 324–327.
4. Priebe, C.E.; Cowen, L.J. A Generalized Wilcoxon–Mann–Whitney statistic. *Comm. Stat. Theory Methods* **1999**, *28*, 2871–2878.
5. Öztürk, Ö. A generalization of Ahmad's class of Mann-Whitney-Wilcoxon statistics. *Aust. N. Z. J. Stat.* **2001**, *43*, 67–74.
6. Read, R.R. The Asymptotic inadmissibility of the sample distribution function. *Ann. Math. Stat.* **1972**, *43*, 89–95.
7. Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **1956**, *27*, 832–837.
8. Parzen, E. On Estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076.
9. Watson, G.S. Smooth regression analysis. *Sankhya Indian J. Stat. Ser. A* **1964**, *26*, 359–372.
10. Nadaraya, E.A. Some new estimates for distribution functions. *Theory Probab. Appl.* **1964**, *9*, 497–500.
11. Yamato, H. Uniform convergence of an estimator of a distribution function. *Bull. Math. Stat.* **1973**, *15*, 69–78.
12. Winter, B.B. Strong uniform consistency of integrals of density estimators. *Can. J. Stat.* **1973**, *1*, 247–253.
13. Winter, B.B. Convergence rate of perturbed empirical distribution functions. *J. Appl. Probab.* **1979**, *16*, 163–173.
14. Reiss, R.D. Nonparametric estimation of smooth distribution functions. *Scand. J. Stat.* **1981**, *8*, 116–119.
15. Falk, M. Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Stat. Neerland.* **1983**, *37*, 73–83.
16. Jones, M.C. The performance of kernel density functions in kernel distribution function estimation. *Stat. Probab. Lett.* **1990**, *9*, 129–132.
17. Swanepoel, J.W.H. Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Comm. Stat. Theory Methods* **1988**, *17*, 3785–3799.
18. Wang, S. Nonparametric estimation of distribution functions. *Metrika* **1991**, *38*, 259–267.
19. Ralescu, S.S. A Remainder estimate for the normal approximation of perturbed sample quantiles. *Stat. Probab. Lett.* **1992**, *14*, 293–298.
20. Shirahata, S.; Chu, I. Integrated squared error of kernel-type estimator of distribution function. *Ann. Inst. Stat. Math.* **1992**, *44*, 579–591.
21. Berg, A.; Politis, D. CDF and survival function estimation with infinite-order kernels. *Electron. J. Stat.* **2009**, *3*, 1436–1454.
22. Leblanc, A. On estimating distribution functions using Bernstein polynomials. *Ann. Inst. Stat. Math.* **2012**, *64*, 919–943.

23. Tenreiro, C. Boundary kernels for distribution function estimation. *REVSTAT-Stat. J.* **2013**, *11*, 169–190.
24. Bouredji, H.; Sayah, A. Bias Correction at End Points in Kernel Density Estimation. In Proceedings of the International Conference on Advances in Applied Mathematics (ICAAM-2018), Sousse, Tunisia, 17–20 December 2018.
25. Oryshchenko, V. Exact mean integrated squared error and bandwidth selection for kernel distribution function estimators. *Comm. Stat. Theory Methods* **2020**, *49*, 1603–1628.
26. Ghorai, J.K.; Susarla, V. Kernel estimation of a smooth distribution function based on censored data. *Metrika* **1990**, *37*, 71–86.
27. Alevizos, F.; Bagkavos, D.; Ioannides, D. Efficient estimation of a distribution function based on censored data. Stat. *Probab. Lett.* **2019**, *145*, 359–364.
28. Ibragimov, I.A.; Linnik, Y.V. *Independent and Stationary Sequences of Random Variables*; Wolters-Noordhoff: Groningen, The Netherlands; 1971; pp. 202–205.
29. Ahmad, I.A. Strong consistency of density estimation by orthogonal series methods for dependent variables with applications. *Ann. Inst. Stat. Math.* **1979**, *31*, 279–288.