The Annals of Statistics
1974, Vol. 2, No. 1, 177-181

ROBUSTNESS OF THE WILCOXON TEST TO A CERTAIN DEPENDENCY BETWEEN SAMPLES

BY MYLES HOLLANDER,[1] GORDON PLEDGER,[2] AND PI-ERH LIN[3]

Florida State University and Stanford University; Florida State University and University of Nebraska-Lincoln; Florida State University

Robustness properties of the two-sample Wilcoxon test are established when the assumption of independence between samples is weakened to allow pairing. In particular, the test is shown to be (asymptotically) conservative when the bivariate distribution governing the pairing is positively quadrant dependent.

1. Introduction and summary. The Wilcoxon test of the hypothesis H_0, that two distribution functions (df's), F_X and F_Y, are equal, is typically based on independent random samples, $X_1, \cdots, X_L$ from $Y_1, \cdots, Y_M$ from F_Y. Serfling (1968) considered robustness of this test to a specific type of dependence *within* samples. The present note establishes properties of the Wilcoxon test when the assumption of independence *between* samples is weakened to allow pairing of X_i and Y_i for some values of i. This type of data arises in various ways:

(i) A researcher may wish to observe the effects (X_i's) of treatment A on L subjects, and at a later date observe the effects (Y_i's) of treatment B on M subjects. However, factors such as a shortage of suitable patients may cause the experimenter to include a few subjects both times.

(ii) Treatments A and B are compared using $2M$ rats from the same litter. Rather than assign the rats to treatments so that each set of M rats has probability $1/\binom{2M}{M}$ of being the "A rats", the researcher blocks (either implicitly or explicitly) on some factor. This blocking may introduce dependencies; however, the blocking is not recorded and the blocks cannot be identified at data analysis time. (We are grateful to Byron W. Brown, Jr. for calling this situation to our attention.)

Received November 1972; revised March 1973.

[1] Research supported in part by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. AFOSR-71-2058 to the Florida State University and National Institutes of General Medical Sciences Grant 5TI-GM 25-15 to Stanford University. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon.

[2] Research supported in part by the Air Force Grant acknowledged in Footnote 1.

[3] Research supported in part by National Institutes of General Medical Sciences Grant 5T 01 GM 913 to the Florida State University.

AMS 1970 *subject classifications*. Primary 62G10, 62G35.

Key words and phrases. Wilcoxon two-sample statistic, robustness, dependence between samples.

(iii) The data may be derived from surveys taken by mail, and it may be impossible to determine whether some respondents contributed to both the X and Y samples.

With this motivation we propose the following model. Let X and Y be random variables with absolutely continuous joint df F and marginals F_X and F_Y. Let $(X_1, Y_1), \cdots, (X_n, Y_n), X_{n+1}, \cdots, X_{n+s}, Y_{n+1}, \cdots, Y_{n+t}$ be independent where (X_i, Y_i) is distributed as (X, Y) for $1 \leqq i \leqq n$, X_{n+i} is distributed as X for $1 \leqq i \leqq s$, and Y_{n+i} is distributed as Y for $1 \leqq i \leqq t$. Note that in example (i) above, n is known and the $(X_1, Y_1), \cdots, (X_n, Y_n)$ pairs can be identified. In example (ii) n may or may not be known but the pairs cannot be identified, and in example (iii) n is not known. The results in Section 2 are apropos to each of these cases.

In the sequel we use the Mann–Whitney form of the two-sample Wilcoxon statistic,

$$U = \sum_{i=1}^{n+s} \sum_{j=1}^{n+t} \varphi(X_i, Y_j)\,, \tag{1}$$

where $\varphi(a, b) = 1$ if $a < b$ and 0 otherwise. In Section 2 we obtain a central limit theorem for U and investigate the effect of the dependence between X_i and $Y_i (1 \leqq i \leqq n)$ on the asymptotic level (asymptotic probability of rejecting H_0 when $F_X = F_Y$) of the test. It is seen that the only effect of F on the asymptotic level is through the parameter

$$\nu(F) =_{\text{def}} \tfrac{1}{2} - \textstyle\int\int F(x, y)\, dF_X(x)\, dF_Y(y)\,. \tag{2}$$

Bounds are obtained for $\nu(F)$, and these bounds lead directly to an assessment of robustness of the Wilcoxon test to the present departure from independence. In addition, the test is shown to be (asymptotically) conservative if F is *positively quadrant dependent*, i.e., if $F(x, y) \geqq F_X(x)F_Y(y)$ for all x, y. This has been shown to be a relatively weak form of positive dependence. See Lehmann (1966) and Esary and Proschan (1972).

2. Conservativeness and robustness. Let $N = s + t + 2n$, and let $\lambda_1, \lambda_2, \lambda_3$ be nonnegative numbers such that $\lambda_1 + \lambda_3 > 0$, $\lambda_2 + \lambda_3 > 0$, and $\lambda_1 + \lambda_2 + 2\lambda_3 = 1$. The notation "$N \rightarrow_* \infty$" will indicate "$N \rightarrow \infty$, $s/N \rightarrow \lambda_1$, $t/N \rightarrow \lambda_2$, and $n/N \rightarrow \lambda_3$." Let $p_1 = \int F_X(x)\, dF_Y(x)$, $p_2 = \int [1 - F_Y(x)]^2\, dF_X(x)$, $p_3 = \int F_X{}^2(x)\, dF_Y(x)$, and define

$$T = \{(n + s)(n + t)/N\}^{\frac{1}{2}} \left[\frac{U}{(n + s)(n + t)} - p_1\right]. \tag{3}$$

THEOREM 1. *As $N \rightarrow_* \infty$, T converges in distribution to a normal random variable with mean* 0 *and variance* $\sigma^2 = (\lambda_2 + \lambda_3)p_2 + (\lambda_1 + \lambda_3)p_3 - p_1^2 - 2\lambda_3\,(\mathrm{Cov}\, F_Y(X_1), F_X(Y_1))$.

SKETCH OF PROOF. The calculation of the asymptotic mean and variance is straightforward. Perhaps the shortest route to the stated form of σ^2 is via the limiting variance of the random variable Z defined below. Asymptotic normality can be shown by a modification of the method used by Hoeffding (1948) and Serfling (1968).

Let $f_1(a) = 1 - F_Y(a), f_2(b) = F_X(b), g(a, b) = [\varphi(a, b) - p_1] - [f_1(a) - p_1] - [f_2(b) - p_1]$, and $Z = (n+s)^{-\frac{1}{2}} \sum_{i=1}^{n+s} [f_1(X_i) - p_1] + (n+s)^{\frac{1}{2}}(n+t)^{-1} \sum_{i=1}^{n+t} [f_2(Y_i) - p_1]$. Asymptotic normality of T will follow if it is shown (i) that Z is asymptotically normal; and (ii) that Z and $(\lambda_2 + \lambda_3)^{-\frac{1}{2}}T$ have the same limiting distribution. To see that (i) holds note that Z has the same limiting distribution as

$$\begin{aligned}
&\lambda_1^{\frac{1}{2}}(\lambda_1 + \lambda_3)^{-\frac{1}{2}}\{s^{-\frac{1}{2}} \textstyle\sum_{i=n+1}^{n+s} [f_1(X_i) - p_1]\} \\
&\quad + [\lambda_2(\lambda_1 + \lambda_3)]^{\frac{1}{2}}(\lambda_2 + \lambda_3)^{-1}\{t^{-\frac{1}{2}} \textstyle\sum_{i=n+1}^{n+t} [f_2(Y_i) - p_1]\} \\
&\quad\quad + n^{-\frac{1}{2}} \textstyle\sum_{i=1}^{n} \{\lambda_3^{\frac{1}{2}}(\lambda_1 + \lambda_3)^{-\frac{1}{2}}[f_1(X_i) - p_1] \\
&\quad\quad\quad + [\lambda_3(\lambda_1 + \lambda_3)]^{\frac{1}{2}}(\lambda_2 + \lambda_3)^{-1}[f_2(Y_i) - p_1]\}
\end{aligned}$$

which is the sum of three independent, properly normalized sums of i.i.d. bounded random variables. To prove that (ii) holds it suffices to show that $E[(\lambda_2 + \lambda_3)^{-\frac{1}{2}}T - Z]^2 \to 0$ as $N \to_* \infty$. Note that asymptotically

$$\begin{aligned}
(4)\qquad &E[(\lambda_2 + \lambda_3)^{-\frac{1}{2}}T - Z]^2 \\
&\doteq (n + s)^{-1}(n + t)^{-2} \textstyle\sum_{i=1}^{n+s} \sum_{j=1}^{n+t} \sum_{u=1}^{n+s} \sum_{v=1}^{n+t} h(i, j, u, v)\,,
\end{aligned}$$

where $h(i, j, u, v) = E[g(X_i, Y_j)g(X_u, Y_v)]$. Because g is bounded and $(n + s)^{-1}(n + t)^{-2}$ is of order N^{-3} as $N \to_* \infty$, we can ignore any k terms of the sum in (4) if k is of order N^2. That the entire sum in (4) is of order N^2 follows from the fact that $E[g(X_i, Y_j) \mid X_i] = E[g(X_i, Y_j) \mid Y_j] = 0$ if $i \neq j$, and consideration of the following cases (where i, j, u, and v represent four distinct indices):

(I) $h(i, j, u, v) = Eg(X_i, Y_j)Eg(X_u, Y_v) = 0$;

(II) $h(i, i, u, v) = h(i, j, u, u) = Eg(X_i, Y_i)Eg(X_u, Y_v) = 0$;

(III) $h(i, j, i, v) = h(i, j, u, j) = E\{E[g(X_i, Y_j)g(X_i, Y_v) \mid X_i]\} = E\{E[g(X_i, Y_j) \mid X_i] \times E[g(X_i, Y_v) \mid X_i]\} = 0$;

(IV) $h(i, j, u, i) = h(i, j, j, v) = E\{E[g(X_i, Y_j) \mid X_i, Y_i]E[g(X_u, Y_i) \mid X_i, Y_i]\} = 0$.

These are the only cases for which the number of terms is larger than $O(N^2)$. □

COROLLARY 1. *If $F_X = F_Y$, then*

$$(5)\qquad \sigma^2 = \tfrac{1}{12} + \lambda_3[2\nu(F) - \tfrac{1}{2}]\,.$$

PROOF. Note that $p_1 = \frac{1}{2}$, $p_2 = p_3 = \frac{1}{3}$, $\text{Cov}\,(F_Y(X_1), F_X(Y_1)) = \text{Cov}\,(F_X(X_1), F_Y(Y_1))$, and, even if $F_X \neq F_Y$,

$$\begin{aligned}
\text{Cov}\,(F_X(X_1), F_Y(Y_1)) &= \textstyle\iint F_X(x)F_Y(y)\, dF(x, y) - \tfrac{1}{4} \\
&= \textstyle\iint F(x, y)\, dF_X(x)\, dF_Y(y) - \tfrac{1}{4} = \tfrac{1}{4} - \nu(F)\,. \qquad \square
\end{aligned}$$

In order to investigate the asymptotic level of the test, we need only consider the case $F_X = F_Y$. (Theorems 2 and 3 include the case $F_X \neq F_Y$, but that case has little relevance in the present context, for then σ^2 is no longer a function of $\nu(F)$.) If $F_X = F_Y$ and the usual independence assumptions hold, then $n = \lambda_3 = 0$ and the asymptotic variance of T is $\frac{1}{12}$. Let $r = 1 + 12\lambda_3[2\nu(F) - \frac{1}{2}]$, the ratio of the asymptotic variance of T in our dependence model to that assuming independence. If z is an approximate α-level critical point [for the one-sided test which rejects for large T values (with p_1 replaced by $\frac{1}{2}$)] obtained via the normal approximation and the independence assumption, then $zr^{\frac{1}{2}}$ is an approximate α-level critical point based on the normal approximation and the present dependence model. Thus for cases where $r \leqq 1$ (and $F_X = F_Y$), the usual Wilcoxon test is (asymptotically) conservative. The next result characterizes the bivariate df's for which $r \leqq 1$.

THEOREM 2. *$r \leqq 1$ if and only if* $\operatorname{Cov}(F_X(X_1), F_Y(Y_1)) \geqq 0$. *In particular, $r \leqq 1$ if F is positively quadrant dependent.*

PROOF. The equivalence follows from the equation $\nu(F) = \frac{1}{4} - \operatorname{Cov}(F_X(X_1), F_Y(Y_1))$ which was established in the proof of Corollary 1. If F is positively quadrant dependent, then $\operatorname{Cov}(F_X(X_1), F_Y(Y_1)) \geqq \iint F_X(x)F_Y(y)\,dF_X(x)\,dF_Y(y) - \frac{1}{4} = 0$. (An alternative proof to the latter statement is by Lemmas 1 and 3 of Lehmann (1966).) □

The following theorem establishes bounds for $\nu(F)$ which imply that $1 - 2\lambda_3 \leqq r \leqq 1 + 2\lambda_3$, thus providing a measure of robustness.

THEOREM 3. $\frac{1}{6} \leqq \nu(F) \leqq \frac{1}{3}$.

PROOF. $|\operatorname{Cov}(F_X(X_1), F_Y(Y_1))| \leqq [\operatorname{Var}(F_X(X_1)) \operatorname{Var}(F_Y(Y_1))]^{\frac{1}{2}}$, and the latter term equals $\frac{1}{12}$ since $F_X(X_1)$ and $F_Y(Y_1)$ are uniformly distributed on $[0, 1]$. □

We now consider the question of sharpness of the bounds provided by Theorem 3. For a given pair of absolutely continuous df's (F_X, F_Y), let $\mathscr{G}$ be the class of all bivariate df's with marginals F_X and F_Y, and let $\mathscr{G}^*$ be the class of all absolutely continuous bivariate df's in $\mathscr{G}$. As we have assumed that F is in $\mathscr{G}^*$ (to avoid the possibility of $X = Y$ ties), the most appropriate (for this paper) sharpness property for bounds on $\nu(F)$ would be that they are sharp for F in $\mathscr{G}^*$. Sharpness in the larger class $\mathscr{G}$ can be seen as follows. Let $G_1(x, y) = \max(0, F_X(x) + F_Y(y) - 1)$ and $G_2(x, y) = \min(F_X(x), F_Y(y))$. Fréchet (1951, 1957), in investigations of the possible behavior of a joint distribution F with given marginals F_X, F_Y, has shown that $G_1(x, y) \leqq F(x, y) \leqq G_2(x, y)$. Both G_1 and G_2 are in $\mathscr{G}$, and a direct calculation yields $\nu(G_1) = \frac{1}{3}$ and $\nu(G_2) = \frac{1}{6}$. The distinction here between $\mathscr{G}$ and $\mathscr{G}^*$ is somewhat academic since, within $\mathscr{G}^*$, $\nu(F)$ can come arbitrarily close to $\frac{1}{6}$ or $\frac{1}{3}$. To see this, let $F(x, y) = \Psi_\rho(\Phi^{-1}(F_X(x)), \Phi^{-1}(F_Y(y)))$, where Ψ_ρ is the bivariate normal df with standard normal marginals Φ and correlation ρ. Then F is absolutely continuous with marginals F_X, F_Y, and $\nu(F) = \frac{1}{4} - \operatorname{Cov}(\Phi(V), \Phi(W))$ where (V, W) have joint df Ψ_ρ. Hence $\nu(F) = \frac{1}{4} + (2\pi)^{-1} \arcsin(-\rho/2)$ and $\nu(F) \to \frac{1}{6}$ or $\frac{1}{3}$ as $\rho \to +1$ or -1.

Better bounds than those of Theorem 3 can be established for a class of bivariate df's with fixed marginals considered by Morgenstern (1956). Let H be a continuous df and, for $-1 \leqq c \leqq 1$, define

$$F(x, y) = H(x)H(y)\{1 + c[1 - H(x)][1 - H(y)]\}\,. \tag{6}$$

Then $F_X = F_Y = H$, F is positively quadrant dependent for $0 \leqq c \leqq 1$, and $\nu(F) = \frac{1}{4} - c/36$ so that $\nu(F)$ is bounded between $\frac{2}{9}$ and $\frac{5}{18}$.

Note that when n is known and the pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$ can be identified, $\nu(F)$ can be consistently estimated from the data. One such consistent estimator $\hat{\nu}$, say, for the special case $F_X = F_Y$, is the proportion of quadruples (X_i, Z_j, Z_k, Y_i) for which the inequalities $X_i < Z_j$ and $Z_k < Y_i$ hold. Here i ranges from 1 to n and both Z_j and Z_k range over the $s + t$ "unpaired" X's and Y's, with $Z_j \neq Z_k$. Then an asymptotically exact test of $F_X = F_Y$, for our dependency model, is obtained by referring $T'/\hat{\sigma}$ to a standard normal distribution, where T' is obtained by replacing p_1 by $\frac{1}{2}$ in (3), and $\hat{\sigma}^2$ is obtained by replacing λ_3 by n/N and $\nu(F)$ by $\hat{\nu}$ in (5).

Acknowledgments. We are grateful to the referee for many improvements and a reference to Fréchet's investigation. Ingram Olkin has called our attention to an article by Whitt (1973) where it is pointed out that the Fréchet bounds were also obtained by Hoeffding (1940).

REFERENCES

ESARY, J. D. and PROSCHAN, F. (1972). Relationships among some concepts of bivariate dependence. *Ann. Math. Statist.* **43** 651-655.

FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A Ser.* 3 **14** 53-77.

FRÉCHET, M. (1957). Sur la distance de deux lois de probabilité. *Publ. Inst. Statist. Univ.* Paris **6**, 183-198.

HOEFFDING, W. (1940). Masstabinvariante Korrelations theorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universitat Berlin* **5** 179-233.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137-1153.

MORGENSTERN, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitt. Math. Statist.* **8** 234-235.

SERFLING, R. J. (1968). The Wilcoxon two-sample statistic on strongly mixing processes. *Ann. Math. Statist.* **39** 1202-1209.

WHITT, W. (1974). Constructing bivariate distributions with extreme correlations. *Ann. Probability* (to appear).

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306