NOTE ON THE WILCOXON-MANN-WHITNEY STATISTIC

By S. K. Zaremba

University College of Swansea

In a note recently published in these Annals, R. F. Potthoff [2], using the bounds for the variance of the Wilcoxon-Mann-Whitney statistic obtained by Z. W. Birnbaum and O. H. Klose [1], attempted to apply this statistic to a test for the coincidence of the medians of two random variables, each of which was assumed to be continuously and symmetrically distributed about its median. The test was claimed to be *consistent for practical purposes*. Unfortunately, consistency for practical purposes was not defined. However, Potthoff's note added some topicality to the question of the possible uses of the Wilcoxon-Mann-Whitney statistic and of *a priori* bounds for its variance.

In an earlier paper [3], I showed that the statistic in question could be used in in a test of the null hypothesis $P[X > Y] = \frac{1}{2}$ when X and Y were *any* random variables, and indeed when the relation "$>$" denoted non-metric preferences, which are relevant to psychology, market research, etc. I further showed that the upper bound obtained by Birnbaum and Klose for the variance of this statistic in the case of continuous distributions still applied under the null hypothesis not only in the case of discontinuous variables, but even in that of non-metric preferences, provided that these should be transitive. This is all we need in connection with the test, but a trivial modification of the proof is sufficient to cover the case when $P[X > Y]$ takes any positive value smaller than 1.

The greatest possible lower bound for the same variance was obtained by me, also for possibly non-metric preferences, under the assumption that the samples of X and Y were of the same size; however, an elaboration of my earlier argument proves that, at least when $P[X > Y] = \frac{1}{2}$, the lower bound obtained by Birnbaum and Klose in the case of continuous distributions applies to the case of non-metric preferences as well, let alone to that of discontinuous distributions. The crux of the argument depends on the following combinatorial lemma:

LEMMA. *Let the* $m + n$ *objects* $x_1, \cdots, x_m, y_1, \cdots, y_n$ ($n \geqq m$; $n - m$ *even*) *be arbitrarily ordered, and let*

$$W = \sum_{i=1}^{m} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \xi_{ik}\xi_{il} + \sum_{k=1}^{n} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \xi_{ik}\xi_{jk},$$

where

$$\begin{aligned} \xi_{ik} &= \tfrac{1}{2} \quad \text{if } x_i \text{ precedes } y_k; \\ &= -\tfrac{1}{2} \quad \text{if } x_i \text{ follows } y_k; \end{aligned}$$

then

$$W \geqq \tfrac{1}{8}m(m-2)(n-1) - \tfrac{1}{24}m(m-1)(m-2). \tag{1}$$

Received 7 August 1964; revised 23 November 1964.

Proof. In the first place, it is easy to verify that if the last x is followed by s y's with $0 \leqq s < \frac{1}{2}(n - m)$, then transferring the last of the y's followed by x's into the place no. $s + 1$ from the end will reduce W. Thus the smallest possible value of W will not be affected by the assumption that the ordering ends with $\frac{1}{2}(n - m)$ y's. Since a reversal of the ordering does not affect W, it follows that we can also assume that this ordering begins with $\frac{1}{2}(n - m)$ y's.

Then the middle $2m$ places of the ordering are occupied by m x's and as many y's, and one proves that W attains its minimum when the x's and the y's alternate. This is best done by induction. We assume that the last $2r$ places among the middle $2m$ are alternately occupied by x's and y's $(0 \leqq r < m)$, the last being, say, an x (the argument would be entirely similar if it were a y). Then, by moving either the last preceding x to the place no. $2r + 1$ from the end, or the last preceding y to the place no. $2r + 2$ from the end, we obtain a new ordering in which the x's and the y's alternate over the last $2r + 2$ places and there is no difficulty in verifying that such a move cannot make W bigger.

Thus W attains its smallest possible value when the ordering begins and ends with $\frac{1}{2}(n - m)$ y's, and consists of an alternation of x's and y's in its middle part. A direct computation of the corresponding value of W completes the proof.

Now let $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ with, say, $n \geqq m$, be a random sample of two populations of objects between which a stochastic relation of transitive preference is well defined (see [3]). Then, if

$$\begin{aligned} u_{ik} &= 1 \qquad \text{when } x_i \text{ is preferred to } y_k\,; \\ &= 0 \qquad \text{when } y_k \text{ is preferred to } x_i\,; \end{aligned}$$

we can form the Wilcoxon-Mann-Whitney statistic

$$V = \sum_{i=1}^{m} \sum_{k=1}^{n} u_{ik}\,.$$

We have

$$\operatorname{var} V = mn[\alpha + (n - 1)\beta + (m - 1)\gamma],$$

where $\alpha = \operatorname{var} u_{ik}$; $\beta = \operatorname{cov}(u_{ik}, u_{il})$ $(k \neq l)$; $\gamma = \operatorname{cov}(u_{ik}, u_{jk})$ $(i \neq j)$. Under the null hypothesis, $E(u_{ik}) = \frac{1}{2}$, and consequently $\alpha = \frac{1}{4}$, $\beta = E[(u_{ik} - \frac{1}{2})(u_{il} - \frac{1}{2})]$, $\gamma = E[(u_{ik} - \frac{1}{2})(u_{jk} - \frac{1}{2})]$, and we can write

$$\text{(2)} \qquad \operatorname{var} V = \tfrac{1}{4}mn + m(m - 1)n[\gamma + (n - 1)(m - 1)^{-1}\beta],$$

which shows that var V is minimized simultaneously with $\gamma + (n - 1)(m - 1)^{-1}\beta$.

Let, in general, μ and ν be any two integers with $0 < \mu \leqq \nu$, and put $m = 2\lambda\mu + 1$, $n = 2\lambda\nu + 1$, where λ is an arbitrary positive integer. If the ordering of the elements considered in the Lemma corresponds, say, to a decreasing order of preference under the stochastic scheme described above, then

$$\gamma + (n - 1)(m - 1)^{-1}\beta = 2[m(m - 1)n]^{-1}E(W),$$

and since $E(W)$ cannot be smaller than the minimum of W found in the Lemma, it follows that

$$\gamma + (n-1)(m-1)^{-1}\beta \geqq \tfrac{1}{4}(m-2)(n-1)[(m-1)n]^{-1} - \tfrac{1}{12}(m-2)n^{-1}$$
$$= (m-2)(n-1)[(m-1)n]^{-1}[\tfrac{1}{4} - \tfrac{1}{12}\mu/\nu].$$

Making λ tend to infinity, we find $\gamma + (\nu/\mu)\beta \geqq \frac{1}{4} - \frac{1}{12}\mu/\nu$, and substituting this in (2), we obtain

$$\text{var } V \geqq mn[\tfrac{1}{4}m - \tfrac{1}{12}(m-1)^2(n-1)^{-1}], \tag{3}$$

which is the value found for the case of continuous distributions by Birnbaum and Klose.

Of course, if ties have a positive probability, as it may happen when we are confronted with two discontinuously distributed random variables, and unless we are satisfied with testing, say $P[X > Y] = \frac{1}{2}$, we have to make $u_{ik} = \frac{1}{2}$ when $x_i = y_k$, and this further diminishes the lower bound of var V, which does not affect the consistency of the test, but lowers its power. However, even (3) shows that when the sizes of the two samples are of different orders of magnitude, then also the upper and lower bounds for var V under the null hypothesis are of different orders of magnitude. It follows that in such cases a test based on *a priori* bounds of var V will have a much lower power than the test originally proposed by me [3] and making use of readily available estimates of β and γ.

REFERENCES

[1] Birnbaum, Z. W. and Klose, O. M. (1957). Bounds for the variance of the Mann-Whitney statistic. *Ann. Math. Statist.* **28** 933–945.

[2] Potthoff, R. F. (1963). Use of the Wilcoxon statistic for a generalized Behrens-Fisher Problem. *Ann. Math. Statist.* **34** 1596–1599.

[3] Zaremba, S. K. (1962). A generalization of Wilcoxon's test. *Monatsh. Math.* **66** 359–370.