ON THE EFFICIENCY OF TWO-SAMPLE MANN-WHITNEY TEST FOR DISCRETE POPULATIONS

By K. C. Chanda[1]

Washington State University

1. Introduction and summary. Most of the nonparametric tests available in the literature use the assumption that the distributions concerned are absolutely continuous. Under this assumption the power efficiencies of these tests for parametric families of distribution are relatively well known. A problem arises, however, when we assume that the distributions are discrete. One would then need to solve the problem of ties, which occur with positive probabilities. One way of getting round this difficulty would be to randomize the order of the tied observations and employ randomized tests. Putter (1955) has presented some interesting results comparing the relative efficiencies of these randomized tests with suitably modified non-randomized tests, when the distributions are discrete. One particular aspect of the nonparametric test, however, has to the best of the author's knowledge, not been discussed so far. This concerns the efficiency of non-parametric tests when the distributions are discrete relative to the most powerful tests available for such distributions when the latter belong to parametric families. One has the impression that the relative efficiencies for such situations are indirectly related to the degree of "discreteness" of the distributions. In other words, the fewer the number of points of probability concentration, the higher is the efficiency. To determine how far this conjecture agrees with the facts, we have discussed the special case of the two-sample Wilcoxon or Mann-Whitney test (Mann and Whitney (1947)). The power efficiency of this test has been worked out for the class of discrete distributions which are of the exponential type, and finally three examples have been discussed.

2. The Mann-Whitney test. Let $(X_1, \ldots, X_m)$ be a sample of m independent observations on a random variable with d.f. $F(x)$ and let $(Y_1, \cdots, Y_n)$ be a sample of n $(\geqq m)$ independent observations on another random variable with d.f. $G(x)$. We assume that both F and G are purely discontinuous, with the same set of discontinuity points which for the sake of simplicity we take as $0, 1, \cdots$. Let $c(u)$ be defined by

$$
\begin{aligned}
c(u) &= 0 && \text{if } u < 0, \\
&= \tfrac{1}{2} && \text{if } u = 0, \qquad (2.1)\\
&= 1 && \text{if } u > 0.
\end{aligned}
$$

Then consider the statistic $w = \sum_{i=1}^{m} \sum_{j=1}^{n} c(x_i - y_j)/mn$. Obviously, w is a

Received April 5, 1962; revised December 13, 1962.

[1] Now at the Iowa State University.

symmetric function of $x_1, \cdots, x_m$ and also symmetric in $y_1, \cdots, y_n$. Then if $m \to \infty$, $m/n \to a$, where a is a finite positive constant, it is well known that $m^{\frac{1}{2}}(W - E(W))$, where $E(W)$ is the expectation of W, has, asymptotically, a normal distribution with zero mean. Thus, if F and G are both absolutely continuous the Mann-Whitney test for the hypothesis $H_0 : F \equiv G$ consists of rejecting H_0 when $|W - E_0(W)|$ is too large, $E_0(W)$ being the expectation of W under H_0. When F and G are purely discontinuous a slight modification in the test-criterion is called for. This is necessary because even under H_0, the variance of W written $V_0(W)$ is not independent of the common unknown d.f. F, as we shall see in Section 3, although $E_0(W)$ is, indeed, independent of F. The new test-critrion W' has been defined in Section 3 and it is proved that under H_0, W' has, asymptotically a normal distribution with zero mean and unit standard deviation. From the point of view of power efficiency it is, perhaps, better to regard this modified Wilcoxon or Mann-Whitney test as a conditional test in the presence of ties, but the cut off points for the test-criterion will then be difficult to determine and the tabulation involved will be prohibitive. It is for this reason we have thought it better to consider only the unconditional sampling properties of W.

3. Asymptotic expressions for the mean and the variance of W. It is easy to show that

$$\begin{aligned} E(W) &= E\{c(X_1 - Y_1)\} = \tfrac{1}{2}P(X_1 = Y_1) + P(X_1 > Y_1) \\ &= \tfrac{1}{2}\sum_{i=0}^{\infty} p_i\bar{p}_i + \sum_{i>j=0}^{\infty} p_i\bar{p}_i = \mu \text{ (say)}, \end{aligned} \tag{3.1}$$

where p_i, $\bar{p}_i$ are the probabilities for F and G respectively at the point $i = 0, 1, \cdots$. Further,

$$\begin{aligned} V\{c(X_i - Y_j)\} &= V\{c(X_1 - Y_1)\} \\ &= \tfrac{1}{4}P(X_1 = Y_1) + P(X_1 > Y_1) - \mu^2 \\ &= \tfrac{1}{4}\sum_{i=0}^{\infty} p_i\bar{p}_i + \sum_{i>j=0}^{\infty} p_i\bar{p}_i - \mu^2 \\ &= A \text{ (say)}. \end{aligned} \tag{3.2}$$

When $i \neq k$,

$$\begin{aligned} &\operatorname{Cov}\{c(X_i - Y_j), c(X_k - Y_j)\} \\ &\quad = \tfrac{1}{4}P(X_i = X_k = Y_j) + \tfrac{1}{2}\{P(X_i > X_k = Y_j) \\ &\qquad + P(X_k > X_i = Y_j)\} + P(X_i, X_k > Y_j) - \mu^2 \\ &\quad = \tfrac{1}{4}\sum_{i=0}^{\infty} p_i^2\bar{p}_i + \sum_{i<j=0}^{\infty} p_i\bar{p}_i p_j + \sum_{i,j>k=0}^{\infty} p_i p_j \bar{p}_k - \mu^2 = B \text{ (say)}. \end{aligned} \tag{3.3}$$

Similarly, when $j \neq k$,

$$\operatorname{Cov}\{c(X_i - Y_j), c(X_i - Y_k)\}$$

$$= \tfrac{1}{4}\sum_{i=0}^{\infty} p_i \bar{p}_i^2 + \sum_{i>j=0}^{\infty} p_i \bar{p}_i \bar{p}_j + \sum_{i>j,k=0}^{\infty} p_i \bar{p}_j \bar{p}_k - \mu^2 = C \text{ (say)}. \tag{3.4}$$

And, finally,

$$\operatorname{Cov}\{c(X_i - Y_j), c(X_k - Y_l)\} = 0 \tag{3.5}$$

for all other values of i, j, k, l. Using results (3.2)–(3.5) we have

$$mV(W) = \sum_{i,k=1}^{m} \sum_{j,l=1}^{n} \operatorname{Cov}\{c(X_i - Y_j), c(X_k - Y_l)\}/mn^2 \tag{3.6}$$

$$= Bm/n + C + o(1).$$

Let, now, $m \to \infty$, $n \to \infty$, $m/n \to a$. Then

$$mV(W) \to aB + C \tag{3.7}$$

and hence $m^{\frac{1}{2}}(W - \mu)/(aB + C)^{\frac{1}{2}}$ has, asymptotically, a normal distribution with zero mean and unit standard deviation. Note that when $F \equiv G$, i.e., $p_i = \bar{p}_i$ for all $i = 0, 1, \cdots$, $\mu = \mu_0 = \frac{1}{2}\sum_{i=0}^{\infty} p_i^2 + \frac{1}{2}(1 - \sum_{i=0}^{\infty} p_i^2) = \frac{1}{2}$, $B = B_0 = (1 - \sum_{i=0}^{\infty} p_i^3)/12$ and $C = C_0 = B_0$. It follows, therefore, that even when $F \equiv G$, the variance of W depends on the common unknown d.f. F. Let us, now, define $\binom{N}{3} Z$ as the number of triplets $(x_i, x_j, x_k)(i \neq j \neq k)$, where for the sake of convenience we have defined $y_i = x_{m+i}$, $1 \leqq i \leqq n$, such that $x_i = x_j = x_k (N = m + n)$. Z, then, can be written in the form $Z = \sum_{i=1}^{M} Z_i/M$, $\left(M = \binom{N}{3}\right)$, where Z_i is a random variable associated with the ith triplet, taking the value 1 if the x's are all equal and 0 otherwise. Considering the fact that the total number of pairs Z_i, Z_j such that there is no x_i common between them is equal to $\binom{N}{3}\binom{N-3}{3}$, the non-zero contribution to $V(Z)$ is of order $\binom{N}{3}\left\{\binom{N}{3} - \binom{N-3}{3}\right\} \Big/ \binom{N}{3}^2$ i.e., of order N^{-1}. Further as $m \to \infty$, $m/n \to a$, $E(Z) \to \sum_{i=0}^{\infty} (\bar{p}_i + ap_i)^3/(1 + a)^3 = \zeta$ (say). It follows, therefore, that Z tends to ζ in mean square and hence in probability. When H_0 is true $\zeta = \sum_{i=0}^{\infty} p_i^3$ and so under H_0, $W' = (12\ mn)^{\frac{1}{2}}(W - \frac{1}{2})/\{(m + n)(1 - Z)\}^{\frac{1}{2}}$ has, asymptotically, a normal distribution with zero mean and unit standard deviation.

Let us now assume that $p_i = p_i(\theta)$ and $\bar{p}_i = p_i(\bar{\theta})$ where θ, $\bar{\theta}$ are unknown values of a parameter occurring in $F(x)$ and $G(x)$. Note that

$$F(x) = \sum_{i \leqq x} p_i(\theta), \qquad G(x) = \sum_{i \leqq x} p_i(\bar{\theta}).$$

Write $\delta = \theta - \bar{\theta}$, $\delta_m = k/m^{\frac{1}{2}}$ where k is an arbitrarily fixed finite positive constant. Let H_m denote the hypothesis $\delta = \delta_m$, let $\mu' = E(W')$ and denote by $V_m(W')$ the variance of W' under H_m. Then it can be shown quite easily that as $m \to \infty$ $m/n \to a$, $Z \to \sum_{i=0}^{\infty} p_i^3$ in probability, $(d\mu'/d\delta)_{\delta=\delta_m}/(d\mu'/d\delta)_{\delta=0} \to 1$, and $V_m(W')/V_0(W') \to 1$. Also, under the sequence of alternatives H_m, W' has, asymptotically, a normal distribution. Define $r(W') = \{(d\mu'/d\delta)_{\delta=0}\}^2/\{mV_0(W')\}$. As $m \to \infty$, $m/n \to a$, $r(W') \to 12D^2/\{(1 + a)(1 - \sum p_i^3)\}$ where $D = \frac{1}{2}\sum_{i=0}^{\infty} p_i\, dp_i(\theta)/d\theta + \sum_{i>j=0}^{\infty} p_i\, dp_j(\theta)\, d\theta$.

4. Asymptotic efficiency of the Mann-Whitney test for the exponential family of distributions. Let $p_x(\theta) = \exp(t\Theta_1 + \Theta + h)$ where t, h are functions of x only and Θ_1, Θ are functions of θ. Assume that Θ_1 is a monotonic function of θ and that the derivatives of Θ_1 and Θ with respect to θ of order two and lower exist and are continuous for all θ. Let

$$T = \left\{\sum_{j=1}^{m} t(X_j) + \sum_{j=1}^{n} t(Y_j)\right\} \Big/ (m + n)^{\frac{1}{2}},$$

$$U = \left\{\sum_{j=1}^{m} t(X_j)/m - \sum_{j=1}^{n} t(Y_j)/n\right\}\left\{mn/(m + n)\right\}^{\frac{1}{2}}.$$

Then the best similar test $\phi = \phi(u, t)$ for $H_0 : \theta = \bar{\theta}$ against $H: \theta \neq \bar{\theta}$ is, by virtue of Theorem 3 Section 4.4 of Lehmann (1959) (after obvious simplification) defined by

$$\begin{aligned} \phi(u, t) &= 1 && \text{when } u < C_1(t) \text{ or } > C_2(t), \\ &= \gamma_i(t) && \text{when } u = C_i(t) \quad i = 1, 2, \\ &= 0 && \text{when } C_1(t) < u < C_2(t), \end{aligned} \tag{4.1}$$

with the C's and γ's determined by $E_0\{\phi(U, T) \mid t\} = \alpha$ and $E_0\{U\phi(U, T) \mid t\} = \alpha E_0(U \mid t)$ where the expectations are calculated under H_0 and conditional on T being fixed at t. These equations can be solved quite easily in view of the fact that the conditional distribution of U given T is independent of θ. Further, one can show, using known properties of symmetric functions that $E_0(U \mid t) = 0$ for almost all t and all values of n.

Note that U and T have, asymptotically, a joint bivariate normal distribution with $\text{Cov}(U, T) = \{V(t(X_1)) - V(t(Y_1))\}\{(mn)^{\frac{1}{2}}/(m + n)\}$. Further

$$\text{Cov}(U, T \mid H_0) = 0$$

and $\text{Cov}(U, T \mid H_m) = O(m^{-\frac{1}{2}})$ so that both for H_0 and the sequence of alternatives $\{H_m\}$, U, T have, asymptotically, independent normal distributions. Then, by a theorem due to Hoeffding and quoted as Theorem 4.1 of Section 7.4 in Fraser (1957) the conditional test $\phi(u, t)$ is, asymptotically, equivalent (both in size and asymptotic power with respect to $\{H_m\}$) to the test $\phi(u)$

defined by

$$\begin{aligned} \phi(u) &= 1 && \text{when } u < C_1 \text{ or } > C_2, \\ &= \gamma_i && \text{when } u = C_i \quad i = 1, 2, \\ &= 0 && \text{when } C_1 < u < C_2, \end{aligned} \tag{4.2}$$

where the C's and γ's are determined by $E_0\{\phi(U)\} = \alpha$ and $E_0\{U\phi(U)\} = 0$. Note that C_1, C_2 both tend to finite limits as $m \to \infty$ and $m/n \to a$ and that the limits have the same magnitude but opposite signs.

Write $\nu = E(U)$, and let $V_0(U)$, $V_m(U)$ denote respectively the variances of U under H_0 and H_m. It is, then easy to check that as $m \to \infty$, $mn \to a$, $(d\nu/d\delta)_{\delta=\delta_m}/(d\nu/d\delta)_{\delta=0} \to 1$, $V_m(U)/V_0(U) \to 1$. The asymptotic normality of U under both H_0 and the sequence H_m has already been established. Define, now $r(U) = \{(d\nu/d\delta)_{\delta=0}\}^2/\{mV_0(U)\}$. Then from Pitman's theorem for two sided tests (as quoted in Fraser (1957), Theorem 3.3 p. 273 and modified for two sided tests) the asymptotic efficiency e of W' relative to U is the limit, as $m \to \infty$ and $m/n \to a$, of the ratio $r(W')/r(U)$.

5. Illustrations.

EXAMPLE 1: *Poisson distribution.* $p_i(\theta) = \exp(-\theta)\theta^i/i!$ $i = 0, 1, \cdots$ and $0 \leqq \theta < \infty$. Note that $dp_i(\theta)/d\theta = p_{i-1}(\theta) - p_i(\theta)$ so that

$$D = -\tfrac{1}{2}\left(\sum_{i=0}^{\infty} p_i p_{i+1} + \sum_{i=0}^{\infty} p_i^2\right).$$

$p_i(\theta)$ is exponential and $e = 3\theta\left(\sum_{i=0}^{\infty} p_i p_{i+1} + \sum_{i=0}^{\infty} p_i^2\right)^2/\left(1 - \sum_{i=0}^{\infty} p_i^3\right)$. It is easy to see that as $\theta \to 0$, $(1 - \sum p_i^3)/\theta \to 3$, $D \to -\frac{1}{2}$ and hence $e \to 1$. Further, we can write $D = -\frac{1}{2}E\{p_X(\theta) + p_{X+1}(\theta)\}$ where X is a Poisson variable. Remembering that $X' = (X - \theta)/\theta^{\frac{1}{2}}$ has, asymptotically (i.e., as $\theta \to \infty$), a normal distribution we can easily show that to order $\theta^{-\frac{1}{2}}$, $p_X(\theta) \sim p_{X+1}(\theta) \sim \exp(-X'^2/2)/(2\pi\theta)^{\frac{1}{2}}$ where $\sim$ denotes equivalence in probability, so that to the same order $E\{p_X(\theta)\} = E\{p_{X+1}(\theta)\} = \{2(\pi\theta)^{\frac{1}{2}}\}^{-1}$. Also, to order θ^{-1}, it can be proved that

$$\sum_{i=0}^{\infty} p_i^3 = E\{p_X^2(\theta)\} = \{2(3)^{\frac{1}{2}}\pi\theta\}^{-1}.$$

TABLE 1

Asymptotic efficiency of the Mann-Whitney test for the Poisson distribution for selected values of θ

θ	e
0.0	1.00
0.2	0.92
0.5	0.91
1.0	0.92
3.0	0.94
∞	0.95

As a result when $\theta \to \infty$, $e \to 3/\pi$. This agrees with the asymptotic efficiency of the Mann-Whitney test for normal distributions (Dantzig (1951)). To order θ^{-1} we can prove that $e = (3/\pi)\{1 - (32\theta)^{-1} + [2(3)^{\frac{1}{2}}\pi\theta]^{-1}\}$. Table 1 gives the values of e for a few selected values of θ.

We now consider two different distributions at the two extreme scales of discreteness.

EXAMPLE 2: *Binomial distribution.* Let X be a binomial variable which can take values 0 and 1 with probablities $1 - \theta$ and θ respectively so that we can write $p_i(\theta) = \theta^i(1 - \theta)^{1-i}$, $i = 0, 1$ and $= 0$ for $i > 1$, $0 \leqq \theta \leqq 1$. Evidently, this is an exponential distribution and it can readily be shown that $e = 1$ in this particular case. Since X can take only two values 0 and 1 W' is, stochastically, equivalent to the standardized difference between the means of the X_i's and the Y_j's which, however, is the optimum test criterion to test the difference between two binomial probabilities θ and $\bar{\theta}$. Hence e should be equal to unity as it, really, is.

EXAMPLE 3: *Geometric distribution.* For this particular case $p_i(\theta) = \theta^i(1 - \theta)$, $i = 0, 1, \cdots, 0 \leqq \theta \leqq 1$. This again is an exponential distribution and it can be shown that $e = 1 - \theta(1 + \theta)^{-2}$. For $\theta = 0$, $e = 1$ and $e \downarrow 0.75$ as $\theta \to 1$.

REFERENCES

DANTZIG, D. VAN (1951). On the consistency and power of Wilcoxon's two sample test. *Indag. Math.* **13** 1–8.

FRASER, D. A. S. (1957) *Nonparametric Methods in Statistics.* Wiley, New York.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

MANN, H. B. AND WHITNEY, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18** 50–60.

PUTTER, J. (1955). The treatment of ties in some nonparametric tests. *Ann. Math. Statist.* **26** 368–386.