Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Sequential-type nonparametric test using Mann-Whitney statistics

Uttam Bandyopadhyay & Atanu Biswas

To cite this article: Uttam Bandyopadhyay & Atanu Biswas (1999) Sequential-type nonparametric test using Mann-Whitney statistics, Journal of Applied Statistics, 26:3, 301-308, DOI: 10.1080/02664769922412

To link to this article: https://doi.org/10.1080/02664769922412

Published online: 02 Aug 2010.

Journal of Applied Statistics, Vol. 26, No. 3, 1999, 301–308

Sequential-type nonparametric test using Mann–Whitney statistics

UTTAM BANDYOPADHYAY[1] **& ATANU BISWAS**[2], [1]*Department of Statistics, University of Calcutta, India and* [2]*Applied Statistics Unit, Indian Statistical Institute, Calcutta, India*

ABSTRACT *The paper provides a nonparametric test for the identity of two continuous univariate distribution functions when observations are drawn in pairs from the populations, by adopting a sampling scheme which, using Mann–Whitney scores, generalizes the existing inverse binomial sampling technique. Some exact performance characteristics of the proposed test are formulated and compared numerically with existing competitors of the proposed test. The applicability of the proposed test is illustrated using real-life data.*

1 Introduction

The problem of testing the identity of two treatment effects has been considered by many authors in recent years (see, for example, Jennison & Turnbull, 1989, 1991, 1993; Kim & DeMets, 1987; Lan & DeMets, 1983; Pocock, 1977), by taking observations in pairs. In the present paper, we consider the problem under a univariate set-up. The problem may be described as follows: Let F_1 and F_2 be two unknown continuous univariate distribution functions. Then, by drawing independent observations from F_1 and F_2 in pairs, our object is to test the null hypothesis

$$H: F_1 = F_2 \tag{1}$$

against a class of one-sided alternatives

$$H_a: F_2 > F_1 \tag{2}$$

when $F_2 > F_1$ means $F_2(x) \geqslant F_1(x)$ for all x with at least one strict inequality. In particular, if we take $F_1 = F(x)$ and $F_2 = F(x - \delta)$, $-\infty < x, \delta < \infty$, or $F_1 = F(x)$ and $F_2 = F(x \exp - (\delta))$, $x > 0$, $-\infty < \delta < \infty$, we have expressions (1) and (2) as

$$H: \delta = 0 \text{ against } H_a: \delta > 0 \tag{3}$$

Correspondence: U. Bandyopadhyay, Department of Statistics, University of Calcutta, 35 Ballygunge Circular Road, Calcutta 700019, India.

0266-4763/99/030301-08 $7.00

For this problem, samples corresponding to F_1 and F_2 may be drawn by adopting the following inverse scheme of sampling: Let X and Y denote the performance of old and new treatments respectively. Suppose that $X \sim F_1$, $Y \sim F_2$ and they are independent. Here, X and Y are observed in pairs. Corresponding to the ith pair (X_i, Y_i), a random variable Z_i is defined by

$$Z_i = \begin{cases} 1, & Y_i > X_i \\ 0, & Y_i \leqslant X_i \end{cases} \tag{4}$$

Obviously, the Z_i terms are iid Bernoulli (p) random variables, where

$$p = p(F, \delta) = \int F_1(x)\, dF_2(x)$$

This equals $\frac{1}{2}$ under H and is greater than $\frac{1}{2}$ under H_a. Now, writing

$$S_n = \sum_{i=1}^{n} Z_i, \qquad n \geqslant 1$$

sampling is stopped at that $n = N$ for which $S_n = r$ is reached for the first time, where r is a predetermined positive integer. If

$$S'_n = \sum_{i=1}^{n} (1 - Z_i), \qquad n \geqslant 1$$

then S'_N has the negative binomial $(r, \frac{1}{2})$ distribution under H. Hence, given $\alpha \in (0, 1)$, we can find a non-negative integer c_α that satisfies

$$P_H\{S'_N \leqslant c_\alpha\} \leqslant \alpha < P_H\{S'_N \leqslant c_\alpha + 1\} \tag{5}$$

Thus, a level α test rule that corresponds to the scheme of sampling mentioned may be as follows: Stop sampling at the nth draw and accept H if

$$S'_n > c_\alpha, \qquad 1 \leqslant n \leqslant N$$

Reject H if

$$S'_N \leqslant c_\alpha \tag{6}$$

This test is unbiased and consistent. The average sample number (ASN) of pairs of the test is smaller than that of the test $[S'_N \leqslant c_\alpha]$. However, power functions of the two tests will remain the same.

For the present problem, Bandyopadhyay & Biswas (1995) suggested some nonparametric group-sequential-type tests using the inverse sampling scheme mentioned here. A natural generalization of this sampling scheme is a scheme which replaces $\{S_n\}$ by Mann–Whitney statistics $\{U_n\}$, and r by $q(r)$, i.e. an integer-valued quadratic function of r. In the next section, a test procedure is proposed using this generalized inverse sampling scheme.

2 Proposed test procedure and its different performance characteristics

Here, the observations are drawn one by one and in pairs. After drawing the ith pair, the differences $Y_j - X_k$ $(1 \leqslant j, k \leqslant i)$ are observed instead of observing the differences $Y_i - X_i$. Then, an indicator variable Z_{ij} is introduced as follows:

$$Z_{ij} = \begin{cases} 1, & Y_j - X_i > 0 \\ 0, & Y_j - X_i \leqslant 0 \end{cases} \tag{7}$$

This leads to the following sequence of Mann–Whitney statistics:

$$U_n = \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}, \qquad n \geqslant 1 \tag{8}$$

The stopping variable N is then defined by

$$N = \min\{n: U_n \geqslant q(r)\} \tag{9}$$

Now, writing

$$U'_n = \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - Z_{ij}), \qquad n \geqslant 1 \tag{10}$$

our proposed test is based on the random sequence

$$\{U'_n, \qquad 1 \leqslant n \leqslant N\} \tag{11}$$

Since U'_n and N are both stochastically smaller under H_a than under H, a level α test is given by expressions (5) and (6) with S'_n replaced by U'_n, $1 \leqslant n \leqslant N$, and c_α replaced by u_α. Finally, to obtain an exactly size α test, the usual randomization technique may be adopted. The test is based on Mann–Whitney statistics, so is obviously exactly distribution free. The test can easily be shown to be unbiased and consistent.

To obtain some exact performance characteristics of the proposed test, we consider, for convenience, simply the non-randomized test. One of the characteristics is measured by the power function of the test. To obtain such a function, we take, for a given F, different δ, and write $\theta = (F, \delta)$. The power function of the test procedure is then

$$P(\theta) = P\{U'_N \leqslant u_\alpha | \theta\} \tag{12}$$

We also consider another characteristic which is the ASN function for the procedure. This can be expressed as

$$A(\theta) = \sum_{n=1}^{\infty} n P_\theta(I_n \mathrm{e}\ \mathcal{J}_n) \tag{13}$$

where the events I_n and $\mathcal{J}_n$ are, respectively, given by

$$I_n = (U_j < q(r), j = 1, 2, \ldots, n-1, U_n \geqslant q(r)) \tag{14}$$

$$\mathcal{J}_n = (U'_j \leqslant u_\alpha, j = 1, 2, \ldots, n-1, U'_n \geqslant u_\alpha) \tag{15}$$

Note that equation (13) has an upper bound $\Sigma\, t p_N(t|\theta)$, where $p_N(t|\theta)$ is the probability mass function (pmf) of N.

Now, to find the distribution of N, we start with the distribution of U_{t_0}, where $t_0 = [(q(r))^{1/2}] + 1$ with $[x]$ as the greatest integer contained in x, and write

$$U_{t_0+1} = U_{t_0} + \bar{U}_{t_0+1} \tag{16}$$

where

$$\bar{U}_{t_0+1} = S_{t_0+1} + S'_{t_0} \tag{17}$$

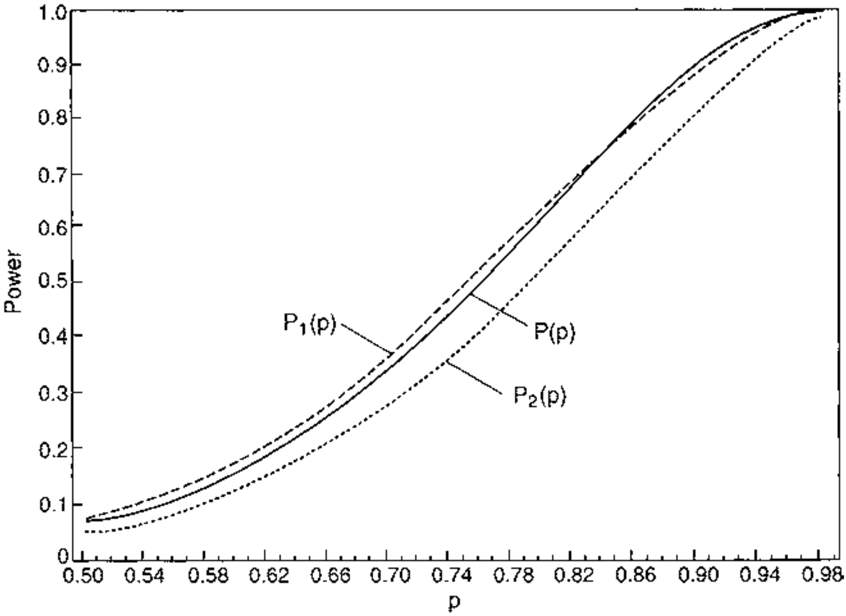

FIG. 1. Power curves for the proposed test and its two competitors.

S_{t_0+1} is the additional score for the inclusion of the variable X_{t_0+1} in the system $\{(X_i, Y_i),\ i = 1, 2, \ldots, t_0\}$, and S'_{t_0+1} is that for the variable Y_{t_0+1} in the system $\{(X_i, Y_i),\ i = 1, 2, \ldots, t_0;\ X_{t_0+1}\}$. Using this breakdown, we can find, for any j, the conditional pmf of U_{j+1}, given $U_j = u_j$. We denote the pmf by $p_{j+1}(u_{j+1} | u_j)$. Finally, we obtain

$$
\begin{aligned}
p_N(t | \theta) &= P\{U_l < q(r),\ l = 1, 2, \ldots, t-1,\ U_t \geqslant q(r) | \theta\} \\
&= \sum_{u_{t_0}=0}^{q(r)-1} \cdots \sum_{u_{t-1}=0}^{q(r)-1} [P_{t_0}(u_{t_0}) p_{t_0+1}(u_{t_0+1} | u_{t_0}) \ldots p_{t-1}(u_{t-1} | u_{t-2}) P\{U_t \geqslant q(r) | u_{t-1}\}]
\end{aligned}
\tag{18}
$$

Finally, in an illustrative computation, we first find the cut-off point u_α, taking $\alpha = 0.065$ and $q(r) = 21$. Then, at different values of θ, we compute the powers and ASNs of the procedure by simulation. These are shown in Figs 1 and 2. The distributions of N at different δ values are shown in Table 1, taking normal, double exponential and exponential parents.

The three values in each cell are those of the three respective parents.

3 Comparison of the proposed test procedure with some existing competitors

In this section, we consider the following competitors of the proposed test procedure:

(i) the test procedure discussed in Section 1;
(ii) fixed-sample-size test, corresponding to the proposed test.

For competitor (i), the power function is given by

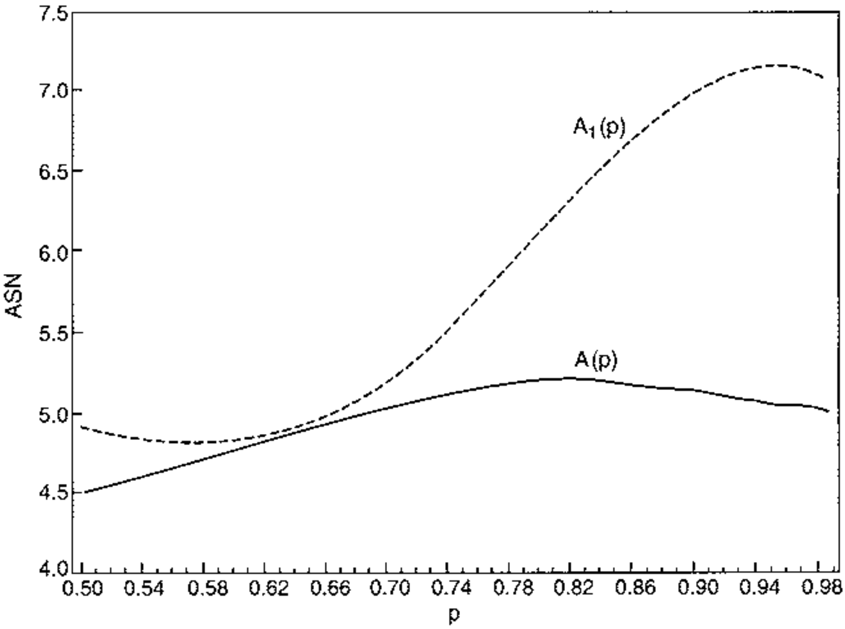

FIG. 2. ASN curves for the proposed test and competitor (i).

$$P_1(\theta) = \sum_{k=0}^{c_\alpha} \binom{r+k-1}{r-1} p^r (1-p)^k \tag{19}$$

To obtain the ASN function, let W_m be a random variable that denotes the number of successes up to m failures. Then, writing

$$\psi(m) = E(W_m \mid W_m < r) = \sum_{s=0}^{r-1} s g(s) \tag{20}$$

where

$$g(s) = \binom{m+s-1}{m-1} (1-p)^m p^s \Bigg/ \sum_{k=0}^{r-1} \binom{m+k-1}{m-1} (1-p)^m p^k \tag{21}$$

we have the expression for the ASN function as

$$A_1(\theta) = \sum_{u=0}^{c_\alpha} (u+r) p_\theta(u) + \sum_{u=c_\alpha+1}^{\alpha} [c_\alpha + 1 + \psi(c_\alpha + 1)] p_\theta(u) \tag{22}$$

where $p_\theta(u)$ is the pmf of the negative binomial (r, p) distribution. Note that r/p is an upper bound of $A_1(\theta)$.

For competitor (ii), a level α test that corresponds to the proposed test is to reject H if

$$U'_{n_0} \leqslant n_0^2 - q(r) \tag{23}$$

where

$$n_0 = \langle [(u_\alpha + q(r))^{1/2}] \rangle \tag{24}$$

TABLE 1. Distributions of N at different δ values, taking normal, double exponential and exponential parents

	$\delta = 0$	$\delta = 0.25$	$\delta = 0.50$	$\delta = 0.75$	$\delta = 1.00$	$\delta = 1.25$	$\delta = 1.50$	$\delta = 1.75$	$\delta = 2.00$	$\delta = 2.25$	$\delta = 2.50$	$\delta = 2.75$	$\delta = 3.00$
$P(N=5)$	0.044	0.103	0.170	0.281	0.397	0.526	0.687	0.774	0.868	0.904	0.968	0.982	0.992
		0.108	0.138	0.222	0.309	0.407	0.475	0.627	0.670	0.730	0.813	0.834	0.876
		0.100	0.151	0.226	0.324	0.454	0.530	0.661	0.736	0.789	0.865	0.897	0.922
$P(N=6)$	0.319	0.421	0.487	0.533	0.487	0.423	0.292	0.216	0.130	0.095	0.032	0.018	0.008
		0.368	0.482	0.537	0.525	0.477	0.459	0.333	0.311	0.263	0.180	0.165	0.123
		0.374	0.467	0.528	0.542	0.468	0.411	0.309	0.251	0.205	0.132	0.103	0.078
$P(N=7)$	0.335	0.324	0.255	0.163	0.106	0.048	0.021	0.010	0.002	0.001	—	—	—
		0.323	0.267	0.186	0.142	0.110	0.064	0.039	0.018	0.006	0.007	0.001	0.001
		0.331	0.275	0.200	0.116	0.070	0.056	0.029	0.012	0.006	0.003		
$P(N=8)$	0.185	0.110	0.066	0.018	0.008	0.003	—	—	—	—			
		0.144	0.088	0.050	0.022	0.006	0.002	0.001	0.001	0.001			
		0.139	0.080	0.037	0.013	0.008	0.003	0.001	0.001				
$P(N=9)$	0.082	0.034	0.020	0.004	0.002								
		0.046	0.019	0.003	0.002								
		0.043	0.021	0.008	0.004								
$P(N=10)$	0.026	0.006	0.002	0.001	—								
		0.010	0.005	0.002	—								
		0.011	0.005	0.001	0.001								
$P(N=11)$	0.007	0.001	—										
		0.001	0.001										
		0.001	0.001										
$P(N=12)$	0.002	0.001											
		—											
		0.001											

TABLE 2. Ovarian weight (grammes) of the soft-shelled turtle *Lissemys punctata punctata*, using four treatments

A_1: Control	A_2: LH-treated	A_3: E-17β-treated	A_4: FSH-treated
28.0	30.0	30.0	50.0
27.4	26.7	26.7	43.4
32.4	27.2	27.2	44.9
25.4	28.0	28.0	45.0
		32.2	51.3

Notes: LH, leutimizing hormone; E-17β, estradiol-17β; FSH, follicle-stimulating hormone. Dose, 15 μg per 100 g body weight for 15 days.

with $\langle [x] \rangle$ as the nearest integer to x. Let us denote the power function of the test by $P_2(\theta)$.

For competitors (i) and (ii), we have to take $r = 7$ and $n_0 = 5$ respectively. Each test is suitably randomized to obtain the size 0.065. Powers and ASNs are computed at different θ values and are shown in Figs 1 and 2. Comparing the graphs, we observe that the proposed test procedure is an improvement, in the sense that it has almost the same power with a smaller ASN.

4 Illustrative example with real-life data

Here, we use an unpublished data set (Sarkar, 1993), which gives the effects of some treatments, including the control treatment, on the ovarian weight of the soft-shelled turtle *Lissemys punctata punctata*. Because the turtle is an endangered animal, from an ethical point of view, the smaller ASN is encouraging. The required portion of the data is given in Table 2. We take $\alpha = 0.065$, $q(r) = 21$ and $u_\alpha = 7$. While comparing A_1 and A_2, we have $U_3 = 2$, $U'_3 = 7$, $U_4 = 7$ and $U'_4 = 9$, which accepts the null hypothesis of equality of treatment effects at a cost of four pairs of sample observations. While comparing A_3 and A_4, we have $U_3 = 9$, $U_4 = 16$ and $U_5 = 25$, yielding $N = 5$. Here, $U'_5 = 0$, which rejects the null hypothesis.

5 Discussion

First, note that, in recent years, the group sequential method has often been adopted for testing the equivalence of two treatments. In this method, a test is carried out in K stages, yielding a maximum of K decision points, where K is a predetermined positive integer. The proposed procedure can be easily extended in this frame as follows.

For the kth (k = 1, 2, ..., K) group, we define

$$N_k = \min\{n: U_n \geqslant q_k(r)\} \tag{25}$$

where one possible value of $q_k(r)$ may be $k^2 q(r)$, since, after kn pairs of (X_i, Y_i) terms, the next n pairs of (X_i, Y_i) terms yield $(2k+1)\, n^2$ possible Z_{ij} terms. Then, defining U'_n as in equation (10), a possible extension would be based on $\{U'_n, 1 \leqslant n \leqslant N_k\}$, $1 \leqslant k \leqslant K$. Now, using these variables sequentially, a test may be described as follows.

Reject H at the kth stage $(1 \leqslant k \leqslant K)$ if

$$U'_{N_l} > c_l, \qquad l = 1, 2, \ldots, k-1, \qquad U'_{N_k} \leqslant c_k$$

Accept H if

$$U'_n > c_k, \text{ for some } n: 1 \leqslant n \leqslant N_k,\ k = 1, 2, \ldots, K \tag{26}$$

where $\{c_1, \ldots, c_k\}$ are non-negative integers determined in such a way that the test has overall level of significance α. Different performance characteristics of the test are under investigation.

Next, we indicate the applicability of the present approach to the following decision-making problem:

$$d_1: \delta = 0, \qquad d_2: \delta > 0, \qquad d_3: \delta < 0 \tag{27}$$

Suppose that Z_{ij}, U_n and U'_n are as in Section 2. Then, we set our stopping variables $N^\star$ as

$$N^\star = \min\{N_1^\star, N_2^\star\} \tag{28}$$

where

$$N_1^\star = \min\{n: U_n \geqslant q(r)\}, \qquad N_2^\star = \min\{n: U'_n \geqslant q(r)\} \tag{29}$$

We stop sampling at the nth draw and accept d_2 or d_3 if $U'_n \leqslant a_\alpha$ or $U_n \leqslant a_\alpha$, and accept d_1 if $U'_{N^\star} > a_\alpha$ and $U_{N^\star} > a_\alpha$. Different performance characteristics of the decision rule can be similarly set, and these are intended to be communicated in a separate paper.

Acknowledgements

The work of the second author was supported by a Research Fellowship from the Council of Scientific and Industrial Research (Sanction 9/28(391)/94-EMR-I).

REFERENCES

BANDYOPADHYAY, U. & BISWAS, A. (1995) Some nonparametric group sequential type tests for two population problems, *Calcutta Statistical Association Bulletin*, 45, pp. 73–91.

JENNISON, C. & TURNBULL, B. W. (1989) Interim analysis: The repeated confidence interval approach (with discussion), *Journal of the Royal Statistical Society, Series B*, 51, pp. 305–361.

JENNISON, C. & TURNBULL, B. W. (1991) Group sequential tests and repeated confidence intervals. *In*: B. K. GHOSH & P. K. SEN (Eds), *Handbook of Sequential Analysis*, pp. 283–311 (New York, Marcel Dekker).

JENNISON, C. & TURNBULL, B. W. (1993) Sequential equivalence testing and repeated confidence with intervals with applications to normal and binary responses, *Biometrics*, 49, pp. 31–43.

KIM, K. & DEMETS, D. L. (1987) Design analysis of group sequential tests based on the type I error spending rate function, *Biometrika*, 74, pp. 149–154.

LAN, K. K. G. & DEMETS, D. L. (1983) Discrete sequential boundaries for clinical trials, *Biometrika*, 70, pp. 649–663.

POCOCK, S. J. (1977) Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, pp. 191–199.

POCOCK, S. J. (1982) Interim analyses for randomized clinical trials: The group sequential approach, *Biometrics*, 38, pp. 153–162.

SARKAR, S. (1993) *Some Aspects of Biology of Female Reproduction in the Soft-shelled Turtle Lissemys Punctata punctata*, Ph.D. Thesis, University of Calcutta, unpublished.