Two-Sample Problem with Bivariate Exchangeable Observations

By L. J. WEI†

George Washington University and University of Michigan, USA

SUMMARY

In comparing two groups of subjects, suppose that each subject has two observations which are probabilistically exchangeable. Univariate rank tests which utilize this symmetry property are proposed for testing the equality of the two groups. Efficiencies of these tests are discussed. Point and interval estimators of the location shift parameter are obtained through the proposed rank tests. Examples are also provided to illustrate the proposed procedures.

1. INTRODUCTION

This paper considers the problem of comparing two groups of subjects when the observation of each subject consists of two probabilistically exchangeable measurements. This type of problem arises often in practice. For example, inversion of the body into an upside-down and totally dependent position is a part of many physical fitness programs. This procedure can relieve back pain and spinal disk pressure. However, it may be harmful to certain groups of patients. Unfortunately, few studies have documented the physiological changes associated with its use. Recently Weinreb *et al.* (1984) have evaluated changes in intraocular pressure in eyes with primary open-angle glaucoma after inversion and compared them with the changes in intraocular pressure in healthy non-glaucomatous eyes. In their study, while the patient was seated, the intraocular pressures of both eyes were measured. Each patient was then inverted and both eyes pressures were taken again. It is reasonable to assume that for each patient the measurement of the pressure in the right eye is probabilistically exchangeable with that in the left eye. Weinreb *et al.* (1984) also confirmed this empirically. Hence, in their analysis, the measurements for the right and the left eyes of each patient were averaged and the usual univariate two-sample tests were then performed using these averages. We will see in this article that this procedure may not be very efficient.

Another interesting study which is similar to the above example can be found in Roy, Gunkel, and Podgor (1986). The aim of this study is to determine whether insulin-dependent diabetic patients, with minimal diabetic retinopathy, have colour vision defects when compared with normal subjects. For each patient, four colour vision tests were performed on both eyes to assess colour vision. Again, the measurement of colour vision in the right eye may be assumed to be probabilistically exchangeable with that in the left eye.

In this paper, we take advantage of the symmetry property in such problems and propose an alternative way to analyse the type of data described above. In Section 2, a class of univariate rank tests is introduced to test the equality of two groups based on the exchangeable paired data. The efficiencies of these tests are also investigated. If one assumes that the intraocular pressure or the measurement of colour vision for the two groups only differ from each other by a location shift parameter θ, then the point and interval estimators of θ can be obtained through the rank tests. This is given in Section 4. The proposed procedures are illustrated using the data from a particular colour vision test conducted by Roy *et al.* (1986). All the results here can be extended to the K-dimensional case.

† *Address for correspondence*: Professor L. J. Wei, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

2. ASYMPTOTICALLY DISTRIBUTION-FREE TESTS

Let (X_{1i}, X_{2i}), $i = 1, \ldots, m$ be i.i.d. random vectors with a continuous distribution function $H(x_1, x_2)$ which is symmetric in its arguments, i.e., $H(x_1, x_2) = H(x_2, x_1)$, for all $(x_1, x_2) \in R^2$. The common marginal distribution function is denoted by F. In the first example of eye studies, X_{1i} is the change in the pressure of the left eye for the ith healthy individual from the seated to the inverted position. The X_{2i} is the change for the right eye. Similarly, let (Y_{1j}, Y_{2j}), $j = 1, \ldots, n$ be the corresponding changes for glaucomatous patients and i.i.d. with a continuous distribution function $K(y_1, y_2)$. The Y's are also assumed to be independent of X's. Again, the K is symmetric with marginal distribution function G.

The hypothesis we wish to test is

$$H_0: H(x, y) = K(x, y), \quad \text{for all} \quad (x, y) \in R^2.$$

Since X_{1i} and X_{2i} are exchangeable for $i = 1, \ldots, m$ as are Y_{1j} and Y_{2j}, $j = 1, \ldots, n$, it seems natural to pool all the X's into one group and pool all the Y's into another and perform a univariate "two-sample" analysis. This leads us to consider the following general linear rank statistic

$$W_N = \frac{1}{2m} \sum_{i=1}^{m} \left[J\left(\frac{R_{1i}}{N}\right) + J\left(\frac{R_{2i}}{N}\right) \right],$$

where $N = 2m + 2n$, J is some score generating funtion defined in the interval $(0, 1)$, and $R_{1i}(R_{2i})$ is the rank of $X_{1i}(X_{2i})$ in the combined sample $\{X_{1i}, X_{2i}, i = 1, \ldots, m; Y_{1j}, Y_{2j}, j = 1, \ldots, n\}$. The choice of the function $J(\cdot)$ has been extensively discussed for the two-sample problem in the literature (cf. Randles and Wolfe, Chapter 9, 1979; Hájek and Šidák, Chapter VI, 1967). For example, in the location problem, if the common density f is known, then the score function $J^*(u) = -f'(F^{-1}(u))/\{f(F^{-1}(u))\}$ produces the asymptotically efficient test.

Note that X_{1i} and X_{2i} are correlated as are Y_{1j} and Y_{2j}, and the test based on W_N is no longer distribution-free. We will show, however, that it is asymptotically distribution-free. The following theorem provides the asymptotic distribution of the standardized W_N for general H and K.

Theorem 1. Let $p_N = m/N$, $q_n = n/N$ and assume that

$$(A1) \quad p_N \to p, \quad q_N \to q, \quad 0 < p, \quad q < \tfrac{1}{2}, \quad \text{as} \quad N \to \infty;$$

$$(A2) \quad |J(s)| \leq c[s(1-s)]^{-(1/2)+\varepsilon}, \quad |J'(s)| \leq c[s(1-s)]^{-(3/2)+\varepsilon},$$

where $0 < \varepsilon < 1$ and C is some constant.

Then, the variable $N^{1/2}(W_N - \mu)$, where $\mu = \int_{-\infty}^{\infty} J(2pF(x) + 2qG(x))dF(x)$, converges in distribution to a normal random variable with mean 0 and variance τ^2, where

$$\tau^2 = q \, \text{var}\left[\int_{-\infty}^{\infty} \{I(Y_{11} \leq y) + I(Y_{21} \leq y)\}J'(2pF(y) + 2qG(y))dF(y)\right] +$$

$$\frac{q}{p^2} \, \text{var}\left[\int_{-\infty}^{\infty} \{I(X_{11} \leq x) + I(X_{21} \leq x)\}J'(2pF(x) + 2qG(x))dG(x)\right],$$

and $I(\cdot)$ is the indicator function. Under H_0, $\mu = \mu_0 = \int_0^1 J(u)du$ and $\tau^2 = \tau_0^2 = q/p\{\int J^2(u)du - 2(\int J(u)du)^2 + \iint J(F(x))J(F(y))dH(x, y)\}$.

Proof. See the Appendix.

Under H_0 and Assumptions (A1) and (A2), the parameter $\eta = \iint J(F(x))J(F(y))dH(x, y)$ can be consistently estimated by

$$\hat{\eta} = 2p_N \iint J(\hat{F}(x))J(\hat{F}(y))d\hat{H}(x, y) + 2q_N \iint J(\hat{G}(x))J(\hat{G}(y))d\hat{K}(x, y)$$

$$= \frac{2}{N} \left\{ \sum_{i=1}^{m} J(\hat{F}(X_{1i}))J(\hat{F}(X_{2i})) + \sum_{j=1}^{n} J(\hat{G}(Y_{1j}))J(\hat{G}(Y_{2j})) \right\},$$

where $\hat{H}$ and $\hat{K}$ are the empirical distribution functions based on $\{(X_{1i}, X_{2i}), i = 1, \ldots, m\}$ nd $\{(Y_{1j}, Y_{2j}), j = 1, \ldots, n\}$, respectively, $\hat{F} = \frac{1}{2}(\hat{F}_1 + \hat{F}_2)$, $\hat{G} = \frac{1}{2}(\hat{G}_1 + \hat{G}_2)$, $\hat{F}_1(x) = \hat{H}(x, \infty)$, $\hat{F}_2(y) = \hat{H}(\infty, y)$, $\hat{G}_1(x) = \hat{K}(x, \infty)$ and $\hat{G}_2(y) = \hat{K}(\infty, y)$.

It follows that under H_0 the variable $N^{1/2}(W_N - \mu_0)/\hat{\tau}_0$ is approximately $N(0, 1)$ for large N, where

$$\hat{\tau}_0^2 = \frac{q_N}{p_N} \left\{ \int J^2(u)du - 2\left(\int J(u)du \right)^2 + \hat{\eta} \right\}.$$

3. EFFICACY OF RANK TEST W_N

It can be seen from the proof of Theorem 1 that the weak convergence of W_N in fact holds uniformly in H and K. This gives us the asymptotic normality for W_N under certain contiguous alternatives.

Firstly, let us evaluate the efficacy of the rank test based on W_N for the location problem. To this end, consider a sequence of alternatives $H_N: K(x, y) = H\left(x + \dfrac{\Delta}{\sqrt{N}}, y + \dfrac{\Delta}{\sqrt{N}} \right)$, for $(x, y) \in R^2$ and a constant Δ. Then, the asymptotic efficacy of the linear rank test W_N is

$$e(W_N) = \frac{2q\left\{ \int \dfrac{dJ(F(x))}{dx} dF(x) \right\}}{\tau_0}$$

for score functions J which satisfy certain smooth conditions (cf. Lehmann, p. 372, 1975). Because of the extra term η in the variance τ_0^2, in general it is not clear that the score function J^* would produce the asymptotically efficient test against H_N. However, if H is a bivariate normal, it is easy to see that the test W_N based on the score J^* or Φ^{-1} is asymptotically efficient, where Φ is the distribution function of $N(0, 1)$. Note that under the normality assumption the maximum likelihood estimator $\hat{\theta}$ of the location shift parameter θ is simply

$$\tfrac{1}{2}[\bar{X}_1 + \bar{X}_2 - \bar{Y}_1 - \bar{Y}_2], \text{ where}$$

$$\bar{X}_i = \sum_{j=1}^{m} X_{ij}/m, \quad \bar{Y}_i = \sum_{j=1}^{n} Y_{ij}/n, \quad i = 1, 2.$$

For the Wilcoxon test, i.e. $J = $ identity function, the asymptotic relative efficiency (ARE) $e(W_N, \hat{\theta})$ of W_N to the test based on $\hat{\theta}$ is $(1 + \rho)/\{(\pi/3) + 2 \sin^{-1}(\rho/2)\}$ when the underlying H is $BN(0, (\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}))$. Table 1 shows that the $e(W_N, \hat{\theta})$ is always greater than $3/\pi$ which is the ARE of the Wilcoxon test to the t-test for the univariate two-sample problem under normality. The ARE of W_N to the procedure T which takes the averages $\tilde{X}_i = (X_{1i} + X_{2i})/2, i = 1, \ldots, m$ and $\tilde{Y}_j = (Y_{1j} + Y_{2j})/2, j = 1, \ldots, n$ first and then applies the two-sample Wilcoxon test to $\tilde{X}$'s and $\tilde{Y}$'s is $(1 + \rho)/\{1 + (6/\pi)\sin^{-1}(\rho/2)\}$. Since we deal with the location problem here, it is no surprise that the gain of efficiency from W_N is small under normality. However, if the underlying distribution F is not normal, then the gain of efficiency for W_N can be substantial. For example, if F is the distribution function of the unit exponential and X_{1i} and X_{2i} are independent, as are Y_{1j} and Y_{2j}, the ARE of W_N to T is 2 for Wilcoxon scores.

Now let us consider the problem of testing for a difference in scale, i.e., the sequence of the alternative hypothesis is now

$$\tilde{H}_N: K(x, y) = H\left(\left(1 + \frac{\Delta}{\sqrt{N}}\right)x, \ \left(1 + \frac{\Delta}{\sqrt{N}}\right)y \right),$$

for $(x, y) \in R^2$ and some Δ. For simplicity, only Klotz's score function, i.e., $J(\cdot) = [\Phi^{-1}(\cdot)]^2$, is considered here (cf. Klotz, 1962). The ARE of W_N to the procedure T with Klotz's score function is $2/(1 + \rho^2)$ when H is the standard bivariate normal with correlation coefficient ρ.

TABLE 1

The asymptotic relative efficiency $100 \times e(W_N, \hat{\theta})$ *under the standard bivariate normal with correlation coefficient* ρ

ρ	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
	95.5	95.9	96.2	96.4	96.6	96.6	96.6	96.5	96.3	95.9	95.5

The test based on W_N is obviously much more powerful than T in this case. The superiority of W_N over T can also be seen for other score generating functions in detecting a difference in scale.

4. ESTIMATION OF THE LOCATION SHIFT PARAMETER

Suppose that the two groups to be compared only differ from each other by an unknown constant θ, i.e., $K(x, y) = H(x + \theta, y + \theta)$. Then the test-based confidence region of θ can be easily obtained by inverting the test W_N (cf. Lehmann, 1963 and Sen, 1966).

Let $W_N = W_N(\mathbf{X}_1, \ldots, \mathbf{X}_m; \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, where $\mathbf{X}_i = (X_{1i}, X_{2i})$ and $\mathbf{Y}_j = (Y_{1j}, Y_{2j})$. Also let $\mathbf{Y}_j + \theta = (Y_{1j} + \theta, Y_{2j} + \theta)$. Then, using Theorem 1 and the fact that, as $N \to \infty$,

$$\iint J(\hat{G}(x))J(\hat{G}(y))d\hat{K}(x, y) \xrightarrow{p} \iint J(G(x))J(G(y))dK(x, y) = \iint J(F(x))J(F(y))dH(x, y),$$

we have that the variable $N^{1/2}\{W_N(\mathbf{X}_1, \ldots, \mathbf{X}_m; \mathbf{Y}_1 + \theta, \ldots, \mathbf{Y}_n + \theta) - \mu_0\}/\hat{\tau}_0 \xrightarrow{d} N(0, 1)$. If $W_N(\theta) = W_N(\mathbf{X}_1, \ldots, \mathbf{X}_m; \mathbf{Y}_1 + \theta, \ldots, \mathbf{Y}_n + \theta)$ is nonincreasing in θ, for large N, an interval $I(\theta)$ of θ can be obtained with any prespecified confidence level.

Now, suppose that

(C1)   large values of W_N reject H_0;

(C2) $W_N(\theta)$ is nonincreasing in θ for every $\mathbf{X}_i$, $i = 1, \ldots, m$ and $\mathbf{Y}_j$, $j = 1, \ldots, n$. Although W_N is not exactly a distribution-free test, the Hodges-Lehmann estimator $\hat{\theta}$ for θ can still be obtained through W_N, where

$$\hat{\theta} = \frac{\theta^* + \theta^{**}}{2}, \; \theta^* = \sup\{\theta: W_N(\theta) > \mu_0\} \text{ and } \theta^{**} = \inf\{\theta: W_N(\theta) < \mu_0\}.$$

By similar arguments given by Hodges and Lehmann (1963) or by Wei and Gail (1983), as $N \to \infty$, the variable $N^{1/2}(\hat{\theta} - \theta)$ converges in distribution to a normal variable with mean 0 and variance $e^{-2}(W_N)$, where $e(W_N)$ is the efficacy of the test based on W_N. Conditions (C1) and (C2) are rather mild. For example, they are satisfied by the Wilcoxon test.

5. AN EXAMPLE

In this section we analyse the results from a particular colour vision test (Farnsworth-Munsell 100-Hue test) conducted by Roy, Gunkel and Podgor (1986) to determine if insulin-dependent diabetic patients have colour vision defects. A high test score indicates that there is a colour vision defect. For each patient test scores were obtained on both eyes. In our analysis there are forty-five insulin-dependent diabetic patients with scores (X_{1i}, X_{2i}) and forty normal patients with score (Y_{1j}, Y_{2j}). The Wilcoxon test $N^{1/2}(W_N - \frac{1}{2})/\hat{\tau}_0$ for testing H_0 yields a value of 2.483 with one-sided p-value .006. This indicates that insulin-dependent diabetic patients, with no-to-minimal diabetic retinopathy had significantly more colour defects. It is interesting to note that if the analyses are done separately for right and left eyes, the corresponding p-values are 0.01 and 0.012, respectively.

If we are willing to assume that the 100-Hue score for the the two groups to be compared

only differ from each other by an unknown constant θ, then the rank estimator based on the Wilcoxon statistics is $\tilde{\theta} = 20.5$. That is, on average, the 100-Hue test score for the diabetic patient is higher than that for the normal control by an amount of 20.5. Furthermore, by inverting the Wilcoxon test, a 90% confidence interval $I(\theta)$ of θ is (6.2, 35.7).

6. REMARKS

The procedures proposed here may be applicable to other medical areas, such as otolaryngology, where correlated replicate observations are obtained from patients. These procedures can be easily generalized to the case where the number of observations contributed by different subjects might not be the same.

Since X_{1i} and X_{2i} are correlated as are Y_{1j} and Y_{2j}, it seems rather difficult, if not impossible, to derive the asymptotically efficient test for testing, for example, a location difference for a given marginal density function f. However, the test based on Wilcoxon score for which J is the identity function for W_N in Section 2, retains both the simplicity and robustness of the univariate two-sample Wilcoxon test. We recommend it for practical use in detecting a possible difference between two groups.

7. APPENDIX

Proof of Theorem 1. We will only sketch the proof here. By an argument similar to that given by Akritas (1984), all the claims made here can be proved rigorously.

Let $\hat{J} = J(p_N \hat{F}_1 + p_N \hat{F}_2 + q_N \hat{G}_1 + q_N \hat{G}_2)$ and $J = J(2pF + 2qG)$. Then $\int \hat{J} d\hat{F}_1 - \int J dF = \int J d(\hat{F}_1 - F) + \int (\hat{J} I_{[X1(1), \infty]} - J) d(\hat{F}_1 - F) + \int (\hat{J} I_{[X1(1), \infty]} - J) dF$, where $X_{1(1)}$ is the smallest order statistic of X_{1i}, $i = 1, \ldots, m$. Integration by parts of the first term gives $-\int (\hat{F}_1 - F) J'(2p_N F + 2q_N G) d(2p_N F + 2q_N G)$. The second term is $o_p(N^{-1/2})$ uniformly in H and K. The third term can be written as

$$\int \{p_N(\hat{F}_1 - F) + p_N(\hat{F}_2 - F) + q_N(\hat{G}_1 - G) + q_N(\hat{G}_2 - G)\} J'(2p_N F + 2q_N G) dF + o_p(N^{-1/2}).$$

If we apply the same kind of approximation to $(\int \hat{J} d\hat{F}_2 - \int J dF)$, then $\tilde{W}_N = W_N - \int J dF = q[\int \{(\hat{G}_1 - G) + (\hat{G}_2 - G)\} J'(2pF + 2qG) dF - \int \{(\hat{F}_1 - F) + (\hat{F}_2 - F)\} J'(2pF + 2qG) dG]$. The $\tilde{W}_N$ is a sum of two sums of independent identically distributed random variables. It follows from the Central Limit Theorem that $N^{1/2} \tilde{W}_N$ is asymptotically normal with mean 0 and the variance specified in the theorem.

Under H_0, note that $\int (I(X_{11} \leqslant x) - F(x)) dJ(F(x)) = -\int J(F(x)) d(I(X_{11} \leqslant x) - F(x)) = -J(F(X_{11})) + \int J(u) du$. Therefore, $\tau_0^2 = (q/2p) \operatorname{var}\{J(F(X_{11})) + J(F(X_{21}))\}$.

REFERENCES

Akritas, M. G. (1984) A simple proof for the Chernoff-Savage Theorem, *Statist. Prob. Letter*, **2**, 39–44.
Hájek, J. and Šidák, Z. (1967) *Theory of Rank Tests*. New York: Academic Press.
Hodges, J. L. and Lehmann, E. L. (1963) Estimates of location based on rank tests. *Ann. Math. Statist.*, **34**, 598–611.
Klotz, J. (1962) Nonparametric tests for scale. *Ann. Math. Statist.*, **33**, 498–512.
Lehmann, E. L. (1963) Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.*, **34**, 1507–1512.
———(1975) *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
Randles, R. H. and Wolfe, D. A. (1979) *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley.
Roy, M. S., Gunkel, R. D., and Podgor, M. J. (1986) Colour vision defects in early diabetic retinopathy. *Archives of Ophthalmol.*, **104**, 225–228.

Sen, P. K. (1966) On a distribution-free method of estimating asymptotic efficiency of a class of non-parametric tests. *Ann. Math. Statist.*, **37**, 1759–1770.
Wei, L. J. and Gail, M. H. (1983) Nonparametric estimation for a scale-change model with censored observations. *J. Amer. Statist. Ass.*, **78**, 382–388.
Weinreb, R. N., Cook, J. and Friberg, T. R. (1984) Effect of inverted body position on intraocular pressure. *Amer. J. Opthalmol.*, **98**, 784–787.