Sankhyā : The Indian Journal of Statistics
Special issue in memory of D. Basu
2002, Volume 64, Series A, Pt. 3, pp 678-692

RE-SAMPLING METHODS FOR THE NON-PARAMETRIC BEHRENS-FISHER PROBLEM

By GUTTI JOGESH BABU
Pennsylvania State University, USA
and
A.R. PADMANABHAN
Monash University, Australia

SUMMARY. A robust solution to the non-parametric Behrens-Fisher problem is proposed, without assuming the symmetry of the underlying distributions. The methodology consists of bootstrapping an appropriately centered version of the Mann-Whitney statistic. A theoretical justification of the bootstrap is also presented.

1. Introduction

The two-sample location problem assumes normality and homogeneity of error variances. However, violations of these assumptions are common (cf. Chow and Liu 1992, Micceri 1989, Wilcox 1987 and 1995). In this paper, we consider the more realistic non-parametric Behrens-Fisher problem of testing for the equality of the medians of two continuous distributions having the same shape, but possibly unequal variances. More precisely, let $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_n)$ be independent samples from respective continuous distributions F_1 and F_2, where

(i) $F_1(x) = F((x - M_X)/\sigma_X)$ and $F_2(y) = F((y - M_Y)/\sigma_Y)$

(ii) F is an arbitrary continuous distribution with median zero and

(iii) σ_X and σ_Y are possibly unequal.

For the standard non-parametric two-sample location problem, when $\sigma_X = \sigma_Y$, the Mann-Whitney test has efficiency of power as well as robust-

Paper received October 2001; revised May 2002.
AMS (1991) *subject classification.* Primary 62G10; secondary 62G35, 62G09.
Keywords and phrases. Bootstrap, Mann-Whitney Statistic.

ness of level (Lehmann, 1975, pp. 76-81). In this case it is also distribution-free and its critical values have been tabulated extensively. However, if the scales are unequal, then this statistic is not even asymptotically distribution-free, since its asymptotic variance would depend on the unknown F. Consequently, using the critical values for equal scales in this case, leads to grossly inflated levels as noted by Fligner and Policello (1981,Table 2, p. 166). Hence they proposed a methodology based on a modification of the Mann-Whitney statistic

$$W = \sum_{i=1}^{m} \sum_{j=1}^{n} I_{\{X_i \leq Y_j\}}$$

when F is symmetric and σ_X and σ_Y are possibly unequal. Under the null hypothesis of equal medians, the symmetry assumption on F implies that $p = P(X_1 \leq Y_1) = 0.5$ and hence $E(W) = 0.5\,mn$ even when the scales σ_X and σ_Y are unequal. This is the case even when the shapes of F_1 and F_2 are different (the generalized Behrens-Fisher problem), as long as they are symmetric. The property that $E(W) = 0.5\,mn$ under symmetry assumption was crucially utilized by Fligner and Policello (1981). They established the asymptotic normality of the studentized version obtained by centering W at $0.5\,mn$ and using an estimate of the asymptotic variance of W.

Nevertheless, most of the distributions arising in Bioavailability studies and Psychology are skewed. See Chow and Liu (1992), Micceri (1989) and Wilcox (1995). Without the symmetry assumption, in general, p differs from 0.5 and depends on the unknown distribution function F. Consequently, in such a case, if W is centered incorrectly at $0.5\,mn$, then the resulting test would be extremely liberal for one tail and conservative for the other (see Remarks 2.1 and 2.2). To overcome this, a methodology without the symmetry assumption is developed in this paper by centering (W/mn) at an estimator $\tilde{p}$ of $p = P(X_1 \leq Y_1)$ and by employing the bootstrap percentile method to obtain critical values. Technical details are given in Section 2. Section 3 provides the simulation studies and a discussion of the results. Section 4 presents an asymptotic theory to justify the use of the bootstrap method. The generalized Behrens-Fisher problem, where F_1 and F_2 could possibly have different shapes, will be considered in a future communication.

2. The Bootstrap Procedures

Let

$$U = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{\{X_i \leq Y_j\}} = \frac{W}{mn}$$

denote the normalized version of the Mann-Whitney statistic, $p = P(X_i \leq Y_j)$, and $\tilde{X}$ and $\tilde{Y}$ denote the sample medians of $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ respectively. For the X-sample, let s_X, $s_{.10,\,X}$ and MADX denote respectively the sample standard deviation, the 10% trimmed sample standard deviation, and median of the absolute deviations from the sample median. Let $s_Y, s_{.10,\,Y}$ and MADY have similar meanings in relation to the Y-sample. Let Hodges-Lehmann (Scale) denote the Hodges-Lehmann type scale estimator given by:

$$med\{(|Y_j - \tilde{Y}|/|X_i - \tilde{X}|) : 1 \leq i \leq m,\ 1 \leq j \leq n\}\,.$$

Bootstrapping U separately from each sample, led to very liberal levels, even when F was normal and $m = n = 20$. So joint bootstrapping was tried after aligning the samples for location and scale. For location – alignment, the obvious choices were the sample medians $\tilde{X}$ and $\tilde{Y}$.

Let Scale (X) and Scale (Y) be the statistics for scale-alignment of the X- and Y- samples respectively. The following choices were studied.

1. Scale $(X) = s_X$, Scale $(Y) = s_Y$
2. Scale $(X) = s_{.10,\,X}$, Scale $(Y) = s_{.10,\,Y}$
3. Scale $(X) =$ MADX, Scale $(Y) =$ MADY
4. Scale $(X) = 1$, Scale $(Y) =$ Hodges-Lehmann (Scale)

We shall present the theory for Case 1 only, as the theory for the other cases is similar. Contrary to conventional wisdom, the best results for U were obtained, when Scale $(X) = s_X$ and Scale $(Y) = s_Y$. Let $(z_1, \ldots, z_{m+n})$ be defined by

$$\begin{aligned} z_i &= (X_i - \tilde{X})/s_X, \quad &\text{if} \quad 1 \leq i \leq m \\ &= (Y_{i-m} - \tilde{Y})/s_Y, \quad &\text{if} \quad m < i \leq m+n\,. \end{aligned}$$

Write

$$Q = m + n, \quad \tilde{p} = \frac{1}{Q^2} \sum_{i=1}^{Q} \sum_{j=1}^{Q} I_{\{z_i\, s_X \leq z_j\, s_Y\}} \quad \text{and} \quad T = \sqrt{n}(U - \tilde{p}).$$

Remark 2.1 If the medians are equal, then under the additional assumption that F_1 and F_2 are symmetric, $p = P(X_1 < Y_1) = 0.5$, even if

F_1 and F_2 have unequal scales or even if F_1 and F_2 have different shapes. However, without symmetry, this is no longer true. If F_1 and F_2 have the same shape, but have unequal scales and are skewed, then p is not known. All that can be said is $p < 0.5$ or > 0.5 depending on (a) the direction of skewness and (b) whether the ratio $\rho = \sigma_Y/\sigma_X$ is > 1 or < 1. For example, if F_1 and F_2 are right-skewed with medians zero and have respective standard deviations 1 and 2, then $p < 0.5$. The inequality would be reversed if either the distributions are left-skewed or the ratio of the scale parameters is < 1. Since we do not assume symmetry, we can work only with $\tilde{p}$, which is an estimate of p.

Let $z_1^*, \ldots, z_Q^*$ be a bootstrap sample from $z_1, \ldots, z_Q$. Two bootstrap procedures are studied.

PROCEDURE I

Let $X_i^* = z_i^*\ s_X,\ i = 1, \ldots, m,\ \ Y_j^* = z_{j+m}^*\ s_Y,\ j = 1, \ldots, n,$

$$U^* = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{\{X_i^* \le Y_j^*\}}\,,$$

s_X^* and s_Y^* denote the sample standard deviations of X^* and Y^*,

$$p^* = \frac{1}{Q^2} \sum_{i=1}^{Q} \sum_{j=1}^{Q} I_{\{z_i^* s_X^* \,\le\, z_j^*\ s_Y^*\}} \quad \text{and} \quad T^* = \sqrt{n}\,(U^* - p^*)\,.$$

PROCEDURE II

Let $X_{i2}^* = z_i^*\ s_X + \tilde{X},\ i = 1, \ldots, m,\ \ Y_{j2}^* = z_{j+m}^*\ s_Y, +\tilde{Y},\ j = 1, \ldots, n,$

$$U_2^* = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{\{X_{i2}^* \le Y_{j2}^*\}}\,,$$

$\tilde{X}^*$ and $\tilde{Y}^*$ denote the sample medians of X_{i2}^*'s and Y_{j2}^*'s,

$$\begin{aligned} v_i &= (X_{i2}^* - \tilde{X}^*)/s_{X^*}, \quad \text{if} \quad i = 1, \ldots, m \\ v_{m+j} &= (Y_{j2}^* - \tilde{Y}^*)/s_{Y^*}, \quad \text{if} \quad j = 1, \ldots, n, \end{aligned}$$

$$p^* = \frac{1}{Q^2} \sum_{i=1}^{Q} \sum_{j=1}^{Q} I_{\{v_i^* s_X^* \,\le\, v_j^*\ s_Y^*\}} \quad \text{and} \quad T_2^* = \sqrt{n}\,(U_2^* - p^*)\,.$$

For Procedure I, T and T^* have the same asymptotic null distribution and this worked well even for small samples (cf. Tables 1-6, Sections 3). For

Procedure II, T and T_2^* have the same asymptotic distribution both under the null hypothesis and under contiguous alternatives. Simulations show that this works reasonably well for fairly large samples (cf. Tables 7 and 8, Section 3). Procedure II is especially useful in the analysis of the limiting distribution under contiguous alternatives. A theory for this is sketched in the section on Asymptotics.

REMARK 2.2 For symmetric F, bootstrapping U itself gives good results. There is no need for centering U. However, for skewed F and unequal scales, this results in tests for one tail being very liberal and the other being very conservative, depending on the direction of skewness and the ratio of scale parameters. Since the symmetry of F cannot be taken for granted, from now on, we shall work only with the centered version $T = \sqrt{n}\ (U - \tilde{p})$.

3. Monte Carlo Simulations

All the results are based on 2000 simulations, involving pseudo-random samples generated by IMSL subroutines, performed on an Alpha 1 computer at Monash University. Let $N(\mu,\ \sigma)$ denote the normal distribution with mean μ and standard deviation σ. The distributions studied were standard normal$= N(0,1)$, 50% $4N = 0.5N(0,1) + 0.5N(0,4)$ and the standard versions of the exponential and the log-normal (centered at their medians).

Two-sided tests at levels $\alpha = 0.025$ and 0.05 were performed. The critical values of T were estimated by the following bootstrap percentile method: From any given sample, 500 bootstrap samples were drawn. Let $t_1^*, t_2^*, \ldots, t_{500}^*$ be the corresponding values of T and $t_{(1)}^*, t_{(2)}^*, \ldots, t_{(500)}^*$ be the corresponding ordered values. Then $t_{(12)}^*, t_{(25)}^*, t_{(475)}^*$ and $t_{(488)}^*$ are respectively the bootstrap estimates of the null 2.5%, 5%, 95% and 97.5% quantiles of T.

Let $C_1 = (-\infty, t_{(12)}^*)$, $C_2 = (-\infty, t_{(25)}^*)$, $C_3 = (t_{(488)}^*, \infty)$ and $C_4 = (t_{(475)}^*, \infty)$. To find the empirical level, the procedure is repeated 2000 times. Let LEFT 2.5%, LEFT 5%, RIGHT 2.5%, and RIGHT 5% denote the percentage of times T falls in C_1, C_2, C_3 and C_4 respectively. Further, let LEVEL 0.025 and LEVEL 0.05 denote the proportions of time T falls in C_3 and C_4 respectively. These values are given in Tables 1-8. Tables 1-6 pertain to Procedure I and 7 and 8 to Procedure II.

TABLE 1. EMPIRICAL LEVELS, $m = n = 20$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEFT	2.5%	2.20	2.75	3.15	2.15	2.25	3.50
LEFT	5%	4.30	5.50	5.35	4.70	4.80	6.40
RIGHT	2.5%	2.55	2.55	3.20	2.60	2.89	2.80
RIGHT	5%	4.55	5.70	6.80	5.00	5.60	5.30
		Exponential			Log-normal		
LEFT	2.5%	2.10	3.09	3.35	2.35	3.55	4.00
LEFT	5%	4.85	6.30	6.40	5.99	7.50	7.10
RIGHT	2.5%	3.50	2.55	2.95	3.55	2.95	3.00
RIGHT	5%	6.00	5.09	5.70	6.25	5.15	5.00

TABLE 2. EMPIRICAL LEVELS, $m = n = 20$ (STUDENTIZED VERSION)

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEFT	2.5%	2.50	2.70	2.00	2.00	2.30	2.60
LEFT	5%	5.20	5.60	4.20	4.75	4.89	5.20
RIGHT	2.5%	2.89	2.40	2.40	2.50	2.30	2.10
RIGHT	5%	5.00	5.70	5.00	4.95	4.50	4.20
		Exponential			Log-normal		
LEFT	2.5%	2.70	3.09	3.20	3.10	4.55	3.60
LEFT	5%	6.30	6.70	6.65	5.35	8.75	7.35
RIGHT	2.5%	2.85	2.50	2.30	3.50	2.45	2.30
RIGHT	5%	6.20	5.05	4.25	6.50	4.59	3.95

REMARK 3.1 The estimator $se(U)$ of the standard error of U given in Fligner and Policello(1981, page 164) is consistent even for skewed F. As $S = \sqrt{n}(U - \tilde{p})/\,se(U) = T/se(U)$ is asymptotically normal, its quantiles can also be obtained from $N(0, 1)$. However, the results are not satisfactory.

The bootstrapped studentized version did not show any significant improvement over the corresponding non-studentized version. For skewed F, especially the log-normal, the results were slightly worse (c.f. Table 2). As a result, we shall not pursue the bootstrap theory for studentized versions. The rest of the results (Tables 3-8) concentrate only on the non-studentized version T.

The empirical levels of the test based on T for sample sizes $m = 20, n = 30$ and $m = n = 30$ are given in Tables 3 and 4 respectively.

TABLE 3. EMPIRICAL LEVELS, $m = 20,\ n = 30$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEFT	2.5%	2.30	2.90	2.90	2.00	2.50	2.35
LEFT	5%	4.60	5.70	6.50	4.50	5.10	5.24
RIGHT	2.5%	2.60	2.50	2.90	2.00	2.90	3.30
RIGHT	5%	5.80	5.00	6.50	4.10	5.20	5.60
		Exponential			Log-normal		
LEFT	2.5%	2.05	2.85	2.75	2.80	2.30	2.90
LEFT	5%	4.19	5.45	6.75	5.10	5.42	5.70
RIGHT	2.5%	3.90	2.60	4.14	4.60	3.60	3.80
RIGHT	5%	6.80	5.50	6.35	7.30	6.40	6.40

TABLE 4. EMPIRICAL LEVELS, $m = n = 30$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEFT	2.5%	2.10	2.30	2.30	2.60	2.00	2.30
LEFT	5%	4.00	5.40	4.40	5.30	4.90	5.40
RIGHT	2.5%	2.20	2.40	2.60	2.89	2.00	2.40
RIGHT	5%	5.40	4.10	5.00	5.70	4.20	4.10
		Exponential			Log-normal		
LEFT	2.5%	3.20	2.30	2.60	3.50	2.80	2.80
LEFT	5%	5.20	5.80	5.50	5.90	5.90	5.50
RIGHT	2.5%	3.30	2.20	2.00	3.60	2.20	2.30
RIGHT	5%	5.80	4.50	4.48	5.90	4.30	4.50

TABLE 5. EMPIRICAL POWERS WHEN SHIFT = 0.5, $m = n = 20$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEVEL	0.025	0.32	0.23	0.156	0.0950	0.092	0.079
LEVEL	0.05	0.43	0.35	0.234	0.1525	0.151	0.138
		Exponential			Log-normal		
LEVEL	0.025	0.55	0.49	0.313	0.38	0.32	0.199
LEVEL	0.05	0.66	0.61	0.416	0.48	0.41	0.284

TABLE 6. EMPIRICAL POWERS WHEN SHIFT = 1.0, $m = n = 20$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEVEL	0.025	0.85	0.68	0.58	0.306	0.222	0.164
LEVEL	0.05	0.92	0.78	0.66	0.408	0.324	0.252
		Exponential			Log-normal		
LEVEL	0.025	0.92	0.91	0.85	0.77	0.73	0.68
LEVEL	0.05	0.96	0.94	0.92	0.85	0.82	0.77

The slight worsening of the empirical levels for skewed distributions, in the case of equal scales (as we go from $m = n = 20$ to $m = 20, n = 30$) seems to be purely an aberration. Asymptotics stabilize at $m = n = 30$ (c.f. Table 4).

Tables 5 and 6 give the powers of the test based on T, when the second sample values are shifted to the right by 0.5 and 1.0 respectively. Because of the shift to the right, the powers are determined essentially by the proportions of rejections corresponding to the right tails.

Procedure II is conservative for small samples but works reasonably well for sample sizes ≥ 50. Table 7 gives the empirical levels, in percentages, for $m = n = 50$. Table 8 gives the empirical powers corresponding to the shift, defined by the 'local' (or 'near') alternative $(50 + 50)^{-1/2} = 0.1$.

Table 7. Empirical Levels, $m = n = 50$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEFT	2.5%	2.30	2.10	1.95	2.35	2.30	1.90
LEFT	5%	4.10	4.20	4.10	4.00	3.90	3.85
RIGHT	2.5%	2.10	2.00	1.80	2.20	2.10	2.00
RIGHT	5%	4.00	3.90	4.00	3.90	3.85	3.80
		Exponential			Log-normal		
LEFT	2.5%	2.30	2.15	1.90	2.30	2.20	1.85
LEFT	5%	4.10	3.95	4.00	3.95	3.50	3.80
RIGHT	2.5%	2.15	2.20	2.10	2.10	2.10	2.00
RIGHT	5%	4.00	3.80	4.15	3.72	3.30	3.70

Table 8. Empirical Powers, $m = n = 50$

		Standard Normal			50% 4N		
σ_Y/σ_X		1.0	1.414	2.0	1.0	1.414	2.0
LEVEL	0.025	0.0380	0.036	0.034	0.0265	0.028	0.0270
LEVEL	0.05	0.0715	0.068	0.063	0.0520	0.050	0.0485
		Exponential			Log-normal		
LEVEL	0.025	0.035	0.038	0.036	0.031	0.029	0.028
LEVEL	0.05	0.076	0.071	0.068	0.054	0.053	0.051

In summary, Tables 1, 3 and 4 show that the Mann-Whitney procedure maintains robustness of level, even for skewed distributions. Table 4 demonstrates that the test performs very well even when $m = n = 30$. Tables 5 and 6 demonstrate the high powers of this procedure when the shifts are 0.5 and 1.0 respectively.

Although, in the case of normal F, the t-test has better asymptotic relative efficiency than the Mann-Whitney test, such an all-round efficiency cannot be achieved, in general, by using the t-test, tests based on sample median, trimmed means or M-estimators. The Mann-Whitney test is more efficient than t-test not only for heavy-tailed F but also for skewed F (cf. Lehmann, 1975, pp. 80-81). The tests based on the sample medians discard all but the middle observations and thus do not perform as well as the Mann-Whitney test when F is normal or skewed (cf. Hajek and Sidak, 1976, p. 278). Tests based on trimmed means or M-estimators respectively discard or downweight the extreme observations. As a consequence, while they are almost on a par with the Mann-Whitney test for symmetric F, they are not as efficient for skewed F.

4. Asymptotics

In this section we shall establish the validity of bootstrap Procedure I for the scales s_X and s_Y under the null hypothesis that the two medians are equal. Similar arguments lead to the validity of the procedure for other scales mentioned in the introduction. Similarly the validity of Procedure II can be established under null hypothesis that the two medians are the same. We shall also sketch the validity arguments for Procedure II under contiguous alternatives.

Recall that M_X and M_Y denote the medians of F_1 and F_2. We assume that $m/n \to \lambda$, where λ is a finite positive number. We start with some notation by defining

$$\begin{aligned} Z_i &= (X_i - M_X)/\sigma_X \quad \text{for} \quad i = 1, \ldots, m, \\ Z_j &= (Y_{j-m} - M_Y)/\sigma_Y \quad \text{for} \quad j = m+1, \ldots, Q, \end{aligned}$$

and $\rho = \sigma_Y/\sigma_X$. Note that $Z_1, \ldots, Z_Q$ are i.i.d. with common distribution F. Let Z be independent of $Z_1, \ldots, Z_Q$ with distribution F. The next theorem pertains to Procedure I.

Theorem 4.1 *Suppose the density f of F satisfies Lipschitz condition of order δ, for some $\delta > 0$ and f has bounded derivative in a neighbourhood of zero. Let $\mathrm{E}(Z^4) < \infty$ and $M_X = M_Y$. For any $0 < \alpha < 1$, if C_α denotes the α-th quantile of T^*, then $P(T \le C_\alpha) - \alpha \to 0$, as $Q \to \infty$.*

The validity of bootstrap Procedure II for contiguous alternatives, when $\sigma_X^{-1}(M_Y - M_X)\sqrt{n} \to \theta$, is established in the next theorem.

Theorem 4.2 *Suppose the density f of F satisfies Lipschitz condition of order δ, for some $\delta > 0$ and f has bounded derivative in a neighbourhood of zero. Let $\mathrm{E}(Z^4) < \infty$ and $\sigma_X^{-1}(M_Y - M_X)\sqrt{n} \to \theta$. For any $0 < \alpha < 1$, if C_α denotes the α-th quantile of T_2^*, then $P(T \le C_\alpha) - \alpha \to 0$, as $Q \to \infty$.*

We require the following lemmas in proving the theorems.

LEMMA 4.1 *Let* $J_1, \ldots, J_m$, *and* $K_1, \ldots, K_n$ *be independent random variables,* J_i*'s identically distributed with distribution* F_J *and* K_j*'s identically distributed with distribution* F_K. *Let* $q = P(J_1 \leq K_1)$, $\bar{F}_K(x) = P(K_1 \geq x)$,

$$R = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (I_{\{J_i \leq K_j\}} - q),$$

$$S = \frac{1}{m} \sum_{i=1}^{m} (\bar{F}_K(J_i) - q) + \frac{1}{n} \sum_{j=1}^{n} (F_J(K_j) - q).$$

Then $mn \ \mathrm{Var}(R - S) \leq 1$.

PROOF. Lehmann (1975, p. 363) shows that $\mathrm{Var}(R - S) = \mathrm{Var}(R) - \mathrm{Var}(S)$. Following the arguments in Lehmann (1975, pp. 363-365), we have,

$$\begin{aligned} mn\,\mathrm{Var}(S) &= n\,\mathrm{Var}(\bar{F}_K(J_1)) + m\,\mathrm{Var}(F_J(K_1)) \\ mn\,\mathrm{Var}(R) &= (n-1)\,\mathrm{Var}(\bar{F}_K(J_1)) + (m-1)\,\mathrm{Var}(F_J(K_1)) + \mathrm{Var}(I_{\{J_1 \leq K_1\}}). \end{aligned}$$

Hence $mn \ \mathrm{Var}(R - S) \leq \mathrm{Var}(I_{\{J_1 \leq K_1\}}) \leq 1$. This completes the proof. □

LEMMA 4.2 *Let* $J_0, J_1, \ldots, J_N$ *be i.i.d. and let* φ *be a measurable function satisfying* $\varphi(x, y) = \varphi(y, x)$, $|\varphi(x, y)| \leq 1$, *and* $\mathrm{E}(\varphi(J_1, J_2)) = 0$. *If*

$$V_N = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} V_{ij},$$

and $S_N = \sum_{i=1}^{N} \mathrm{E}(V_N | J_i)$, *where* $V_{ij} = \varphi(J_i, J_j)$. *Then* $N(N-1) \ \mathrm{Var}(V_N - S_N) \leq 2$.

PROOF. Since $S_N = 2N^{-1} \sum_{i=1}^{N} \mathrm{E}(V_{0i} | J_i)$, a simple algebra leads to

$$(N(N-1)/2)\ \mathrm{Var}(V_N) = \mathrm{Var}(V_{12}) + 2(N-2)\,\sigma^2,$$

$$\mathrm{Var}(V_N - S_N) = \mathrm{Var}(V_N) - \mathrm{Var}(S_N) \quad \text{and} \quad N\ \mathrm{Var}(S_N) = 4\sigma^2,$$

where $\sigma^2 = \mathrm{Cov}(V_{12}, V_{23}) = \mathrm{Var}(\mathrm{E}(V_{12} | J_1))$. Hence

$$N(N-1)\ \mathrm{Var}(V_N - S_N) = 2\,\mathrm{Var}(V_{12}) - 4\,\sigma^2 \leq 2,$$

which completes the proof. □

PROOF OF THEOREM 4.1. Define $\tilde{\rho} = s_X/s_Y$. Let $\tilde{Z}_1$, $\tilde{Z}_2$ denote the sample medians of $Z_1, \ldots, Z_m$ and $Z_{m+1}, \ldots, Z_Q$ respectively. Observe that

$$\begin{aligned} z_i &= \sigma_X(Z_i - \tilde{Z}_1)/s_X, && \text{if} \quad 1 \le i \le m \\ &= \sigma_Y(Z_{i-m} - \tilde{Z}_2)/s_Y, && \text{if} \quad m < i \le Q, \end{aligned}$$

and $z_i s_X \le z_i s_Y$ if and only if

$$Z_i \le \begin{cases} \rho(Z_j - \tilde{Z}_2) + \tilde{Z}_1, & \text{if} \quad i \le m < j \\ \tilde{\rho}(Z_j - \tilde{Z}_1) + \tilde{Z}_1, & \text{if} \quad i, j \le m \\ \tilde{\rho}(Z_j - \tilde{Z}_2) + \tilde{Z}_2, & \text{if} \quad i, j > m \\ \rho^{-1}\tilde{\rho}^2(Z_j - \tilde{Z}_1) + \tilde{Z}_2, & \text{if} \quad j \le m < i\,. \end{cases}$$

If $\hat{F}$ and $\hat{G}$ denote the empirical distributions of $Z_1, \ldots, Z_m$ and $Z_{m+1}, \ldots, Z_Q$ respectively, then

$$\begin{aligned} Q^2\tilde{p} \;=\; & m\sum_{j=1}^{m} \hat{F}(\tilde{\rho}(Z_j - \tilde{Z}_1) + \tilde{Z}_1) + m \sum_{j=m+1}^{Q} \hat{F}(\rho(Z_j - \tilde{Z}_2) + \tilde{Z}_1) \\ & + n\sum_{j=1}^{m} \hat{G}(\rho^{-1}\tilde{\rho}^2(Z_j - \tilde{Z}_1) + \tilde{Z}_2) + n \sum_{j=m+1}^{Q} \hat{G}(\tilde{\rho}(Z_j - \tilde{Z}_2) + \tilde{Z}_2). \qquad (1) \end{aligned}$$

Since the median of F is zero and $\mathrm{E}(Z^4) < \infty$, it is well known that,

$$\tilde{Z}_k = O_p(Q^{-1/2} \log\log Q) \;\text{ for }\; k = 1, 2, \quad \text{and} \quad \tilde{\rho} - \rho = O_p(Q^{-1/2} \log\log Q).$$

As $M_X = M_Y$, clearly $p = p(Z_1 \le \rho Z)$. If

$$\varphi(x, y) = \frac{1}{2}(I_{\{x \le \rho y\}} + I_{\{y \le \rho x\}} - 2p), \quad \text{and} \quad U_0 = Q^{-2} \sum_{1 \le i, j \le Q} I_{\{Z_i \le \rho Z_j\}}$$

then $|\varphi(x, y)| \le 1$ and

$$\begin{aligned} & |U_0 - p - \frac{2}{Q(Q-1)} \sum_{1 \le i < j \le Q} \varphi(Z_i, Z_j)| \\ & \le \; |U_0 - p - \frac{2}{Q^2} \sum_{1 \le i < j \le Q} \varphi(Z_i, Z_j)| + \frac{1}{Q} \\ & \le \; \frac{1}{Q^2} \sum_{i=1}^{Q} |I_{\{Z_i \le \rho Z_i\}} - p| + \frac{1}{Q} \le \frac{2}{Q}. \end{aligned}$$

Now by Lemma 4.2, it follows that

$$Q(Q-1)\mathrm{Var}\left(\frac{2}{Q(Q-1)}\sum_{1\le i<j\le Q}\varphi(Z_i,Z_j) - \frac{1}{Q}\sum_{i=1}^{Q}(F(\rho Z_i)+1-F(\rho^{-1}Z_i)-2p)\right)\le 2.$$

So

$$U_0 - p - \frac{1}{Q}\sum_{i=1}^{Q}(F(\rho Z_i)+1-F(\rho^{-1}Z_i)-2p) = o_p(Q^{-1/2}) \tag{2}$$

uniformly in F.

Now note that as the density of F satisfies Lipschitz condition of order $\delta > 0$, by (1) and the arguments following (2.14) of Babu and Singh (1978), and by the laws of large numbers we obtain

$$\begin{aligned}
Q^2(\tilde{p}-U_0) &= m\sum_{j=1}^{m}(F(\tilde{\rho}(Z_j-\tilde{Z}_1)+\tilde{Z}_1)-F(\rho Z_j)) \\
&\quad +m\sum_{j=m+1}^{Q}(F(\rho(Z_j-\tilde{Z}_2)+\tilde{Z}_1)-F(\rho Z_j)) \\
&\quad +n\sum_{j=1}^{m}(F(\rho^{-1}\tilde{\rho}^2(Z_j-\tilde{Z}_1)+\tilde{Z}_2)-F(\rho Z_j)) \\
&\quad +n\sum_{j=m+1}^{Q}(F(\tilde{\rho}(Z_j-\tilde{Z}_2)+\tilde{Z}_2)-F(\rho Z_j)) + o_p(Q^{3/2}) \\
&= m^2(\tilde{\rho}-\rho)\mathrm{E}(Zf(\rho Z)) + m^2(1-\rho)\tilde{Z}_1\mathrm{E}(f(\rho Z)) \\
&\quad +n^2(\tilde{\rho}^2-\rho^2)\rho^{-1}\mathrm{E}(Zf(\rho Z)) + n^2(\tilde{Z}_2-\rho\tilde{Z}_1)\mathrm{E}(f(\rho Z)) \\
&\quad +mn(\tilde{Z}_1-\rho\tilde{Z}_2)\mathrm{E}(f(\rho Z)) + mn(\tilde{\rho}-\rho)\mathrm{E}(Zf(\rho Z)) \\
&\quad +mn(1-\rho)\tilde{Z}_2\mathrm{E}(f(\rho Z)) + o_p(Q^{3/2}) \\
&= (\tilde{\rho}-\rho)\mathrm{E}(Zf(\rho Z))(m^2+mn+2n^2) \\
&\quad +\mathrm{E}(f(\rho Z))\tilde{Z}_1(m^2(1-\rho)+mn-\rho n^2) \\
&\quad +\mathrm{E}(f(\rho Z))\tilde{Z}_2(mn(1-2\rho)+n^2) + o_p(Q^{3/2}).
\end{aligned} \tag{3}$$

Clearly, $\mathrm{E}(f(\rho Z)) = g(0)$, where g is the density of $Z_1 - Z\rho$. Also, observe that $f(\rho Z) \le a|Z|^\delta$ for some $a > 0$. As $\frac{m}{n} \to \lambda$, we have by Bahadur's representation of quantiles (Babu and Singh, 1978), Lemma 4.1 and (2)–(3),

$$U - \tilde{p} = (U-p) - (U_0-p) - (\tilde{p}-U_0)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\ell_1(Z_i) + \frac{1}{n}\sum_{i=m+1}^{Q}\ell_2(Z_i) + o_p(Q^{-1/2}),$$

where $\mathrm{E}(\ell_j(Z))^2 < \infty$ for $j = 1, 2$. Thus T is asymptotically distributed as normal with mean zero and variance τ^2, for some $\tau > 0$. Essentially the same steps as above and the proof of Theorem 5 of Babu and Singh (1984) lead to the conclusion that for almost all sample sequences, T^* is asymptotically distributed as normal with mean zero and variance τ^2. The details are omitted. The theorem now follows from Lemma 2.1 of Babu and Bose (1988).

Proof of Theorem 4.2. Note that

$$\begin{aligned} F_1(Y_j) - p &= F(\rho Z_j + \theta_n) - P(Z_1 - \rho Z \le \theta_n) \\ &= F(\rho Z_j + \theta_n) - F(\rho Z_j) + F(\rho Z_j) - P(Z_1 \le \rho Z) \\ &\quad -(P(Z_1 - \rho Z \le \theta_n) - P(Z_1 - \rho Z \le 0)), \end{aligned}$$

where $\theta_n = \sigma_X^{-1}(M_Y - M_X)$. Clearly,

$$\begin{aligned} \mathrm{Var}(F(\rho Z_j + \theta_n) - F(\rho Z_j)) &\le \mathrm{Var}(I_{\{Z-\rho Z_j \le \theta_n\}} - I_{\{Z-\rho Z_j \le 0\}}) \\ &\le |P(Z_1 - \rho Z \le \theta_n) - P(Z_1 - \rho Z \le 0)|. \end{aligned}$$

Since by assumption $\theta_n\sqrt{n} \to \theta$, we have

$$\begin{aligned} \sqrt{n}(P(Z_1 - \rho Z \le \theta_n) - P(Z_1 - \rho Z \le 0)) &= \sqrt{n}\mathrm{E}(F(\rho Z + \theta_n) - F(\rho Z)) \\ &\to \theta\mathrm{E}(f(\rho Z)). \end{aligned}$$

Thus as in Theorem 4.1,

$$U - \tilde{p} = (U - p) - (U_0 - P(Z_1 \le \rho Z)) - (\tilde{p} - U_0) + (p - P(Z_1 \le \rho Z)),$$

leads to the asymptotic normality of T with mean $\theta\mathrm{E}(f(\rho Z))$ and variance τ^2. Similarly, the bootstrapped version T_2^* has also the same asymptotic distribution. The rest of the proof is similar to that of Theorem 4.1.

Acknowledgments. We are deeply grateful to the referee for valuable suggestions and comments, to Professor T.P. Hettmansperger of Department of Statistics at Penn State University for useful conversations and to Professors H. Keselman and A. Othman of Department of Psychology at University of Manitoba for having provided copies of reprints of the papers mentioned in the introduction. The work of G.J. Babu is supported in part by NSF grant DMS-0101360.

References

Babu, G.J. and Bose, A. (1988). Bootstrap confidence intervals, *Statist. Probab. Lett.*, **7**, 151-160.

Babu, G.J. and Singh, K. (1978). On probabilities of moderate deviations for dependent processes, *Sankhyā Ser. A*, **40**, 28-37.

Babu, G.J. and Singh, K. (1984). Asymptotic representations related to jackknifing and bootstrapping L-statistics, *Sankhyā Ser. A*, **46**, 195-206.

Chow, S.C. and Liu, J.P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies.* Marcel Dekker, New York.

Fligner, M.A. and Policello, G.E. (1981). Robust Rank Procedures for the Behrens-Fisher Problem, *J. Amer. Statist. Assoc.*, **76**, 162-178.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests.* Academic, New York.

Lehmann, E.L. (1975). *Nonparametrics: statistical methods based on ranks*. Holden-Day, San Francisco, CA; McGraw-Hill International, New York-Düsseldorf.

Micceri, T. (1989) The unicorn, the normal curve and other improbable creatures, *Psychological Bull.*, **105**, 156-166.

Wilcox, R.R. (1987) New designs in analysis of variances, *Annual Rev. Psychology*, **38**, 29-60.

Wilcox, R.R. (1995). ANOVA: The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies, *British J. Math. Statist. Psychol.*, **49**, 163-170.

Gutti Jogesh Babu
Department of Statistics
Pennsylvania State University
University Park
PA 16802, USA
E-mail: babu@stat.psu.edu

A.R. Padmanabhan
Department of Maths. and Stats.
Monash University
P.O. Box 28M
Victoria 3800, Australia
E-mail: appaswany.padmanabhan@sci.monash.edu.au
padmanabhan63@hotmail.com