THE INSTITUTE OF STATISTICS

RESAMPLING FROM CENTERED DATA IN THE TWO-SAMPLE PROBLEM

D. Boos[1], P. Janssen[2] and N. Veraverbeke[2]

[1]North Carolina State University, Raleigh, N.C. 27695-8203 U.S.A.
[2]Limburgs Universitair Centrum, Diepenbeek, Belgium

Institute of Statistics Mimeo Series No. 1901

September, 1987

Research supported by the NFWO (National Scientific Foundation), Belgium

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

RESAMPLING FROM CENTERED DATA IN THE TWO-SAMPLE PROBLEM

D. Boos[1], P. Janssen[2] and N. Veraverbeke[2]

[1]North Carolina State University, Raleigh, N.C. 27695-8203 U.S.A.
[2]Limburgs Universitair Centrum, Diepenbeek, Belgium

ABSTRACT

Bootstrap and permutation approximations to the distribution of U-statistics are shown to be valid when the resampling is from residuals in the two-sample problem. The motivation for using residuals comes from testing for homogeneity of scale in the presence of nuisance location parameters. New asymptotic results for U-statistics with estimated parameters are key tools in the proofs.

Key words and phrases. Bootstrap, permutation, resampling, location alignment, scale parameters, U- and V-statistics, asymptotic theory, nuisance parameters.

AMS 1980 subject classification. Primary 62G10; secondary 62E20.

I. INTRODUCTION

Suppose $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ are independent samples from distribution functions F_1 and F_2 and that we want to test H_0: $\theta_1 = \theta_2$ versus H_a: $\theta_1 > \theta_2$, where θ_i is some scale or variance parameter of F_i. Let $\hat{\theta}_1 = \hat{\theta}_1(X_1,\ldots,X_m)$ be an estimator for θ_1 and let $\hat{\theta}_2 = \hat{\theta}_2(Y_1,\ldots,Y_n)$ be an estimator for θ_2. If for example θ_i is the variance of F_i and $\hat{\theta}_i$ is the sample variance, Box (1953) and many others have firmly established that the classical normal likelihood ratio statistic $\hat{\theta}_1/\hat{\theta}_2$, used with F(m-1, n-1) critical values, gives a test which is very sensitive to the normal assumption. Therefore, when normality is suspect it is natural to consider resampling methods for choosing critical values.

In this paper we deal with the problem of obtaining bootstrap and permutation critical values for a general test statistic of the form

$$(1.1) \qquad T_{mn} = \left(\frac{mn}{m+n}\right)^{1/2} [\log \hat{\theta}_1 - \log \hat{\theta}_2] \ ,$$

where $\hat{\theta}_i$ is of the form $\theta(\hat{F}_i)$, with $\theta(\cdot)$ a von Mises functional and $\hat{F}_1$ and $\hat{F}_2$ are the empirical distribution functions of the two samples. From an appropriate resampling plan we take a large number of resamples and compute T^*_{mn} (the resampled value of T_{mn}) and reject H_0 if T_{mn} is larger than the $(1-\alpha)$ quantile of the empirical distribution function of the T^*_{mn}.

Naive intuition might suggest that one obtain critical values via resampling with replacement (bootstrap) or without replacement (permutation) from

(I) the combined sample $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$

or by bootstrap resampling from

(II) the samples $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ separately.

Assume that c^*_I is the critical point obtained from resampling plan (I). The following example shows that the level of a nominal level α test based on

T_{mn} with c_I^* as critical value need not converge to α : i.e., under H_0 we have $P(T_{mn} > c_I^*) \not\to \alpha$ as $\min(m,n) \to \infty$. Hence (I) is an unacceptable procedure.

EXAMPLE. Suppose that we want to test H_0: $\theta_1 = \theta_2$ versus H_a: $\theta_1 > \theta_2$ in the following semi-parametric framework: $F_1(x) = F_0((x-\mu_1)/\sigma_1)$, $F_2(x) = F_0((x-\mu_2)/\sigma_2)$ with $\sigma_i^2 = \theta_i$, unknown F_0 and unknown location parameters μ_1 and μ_2. If F_0 has zero mean and unit variance, then μ_i is the mean of F_i and σ_i^2 is the variance of F_i. Consider the statistic given by (1.1) with $\hat{\theta}_1$ and $\hat{\theta}_2$ the sample variances. Suppose F_0 has a finite fourth moment. It is easy to derive that for $m,n \to \infty$ with $m/(m+n) \to \lambda \in (0,1)$, T_{mn} has a limiting distribution which under H_0 is given by $N(0; \beta_2(F_0) - 1)$, where for any distribution function K we denote the kurtosis by

$$(1.2) \qquad \beta_2(K) = \left[\int(x - \int x dK(x))^4 dK(x)\right] \Big/ \left[\int(x - \int x dK(x))^2 dK(x)\right]^2 .$$

From Bickel and Freedman (1981) it also follows that the bootstrapped quantity T_{mn}^* (obtained via resampling with replacement from the combined sample) has a conditional limit distribution which, under H_0, is $N(0; \beta_2(G) - 1)$, where $G(x) = \lambda F_1(x) + (1-\lambda)F_2(x)$. Therefore the $(1-\alpha)$ quantile of the empirical distribution function of the T_{mn}^* will converge almost surely to $z_{1-\alpha}(\beta_2(G) - 1)^{1/2}$, where $z_{1-\alpha}$ is the standard normal $(1-\alpha)$ quantile. Since $\beta_2(G)$ can be quite different from $\beta_2(F_0)$, a nominal $\alpha = 0.05$ test will have a different asymptotic significance level. Take for example F_0 standard normal, $\lambda = 1/2$, $\theta_1 = \theta_2 = 1$ and $\mu_1 - \mu_2 = 3$ and use the fact that the mean of G is $\lambda\mu_1 + (1 - \lambda)\mu_2$ and, under H_0, the variance of G is $\theta_1 + \lambda(1-\lambda)(\mu_1-\mu_2)^2$, to obtain $\beta_2(G) = 2.04 \neq \beta_2(F_0) = 3$. It turns out that the level of a nominal $\alpha = 0.05$ test based on T_{mn} with T_{mn}^* critical values converges to 0.12.

Let c_{II}^* be the critical point obtained from plan (II). Combining results in Bickel and Freedman (1981) and the ideas in Section 2 of this paper, it is not hard to show that $P(T_{mn} > c^*_{II}) \to \alpha$ under both H_0 and H_a; so also (II) is unacceptable.

A possible solution to these problems is to adjust for the location differences by centering the samples before resampling, i.e., by resampling from

(III) the location aligned and then combined sample $X_1-\hat{\mu}_1,\ldots,X_m-\hat{\mu}_1$, $Y_1-\hat{\mu}_2,\ldots,Y_n-\hat{\mu}_2$, where $\hat{\mu}_i$ is an estimator for the location parameter μ_i of F_i, $i = 1,2$.

This means that both resamples are drawn from the same distribution $H_{mn}(x) = (m/(m+n))\hat{F}_1(x+\hat{\mu}_1) + (n/m+n))\hat{F}_2(x+\hat{\mu}_2)$. For example, when comparing variances, we will resample from $X_1 - \bar{X},\ldots,X_m - \bar{X}$, $Y_1 - \bar{Y},\ldots,Y_n - \bar{Y}$ where $\bar{X}$ and $\bar{Y}$ are the sample means of the two samples. Results in Sections 2 and 3 show that the $(1-\alpha)$ quantile obtained from the empirical distribution function of the resampled values converges almost surely to $z_{1-\alpha}(\beta_2(H) - 1)^{1/2}$, where $H(x) = \lambda F_1(x + \mu_1) + (1 - \lambda)F_2(x + \mu_2) = \lambda F_0(x/\sigma_1) + (1 - \lambda)F_0(x/\sigma_2)$, a distribution function which, under H_0, has the correct kurtosis $\beta_2(H) = \beta_2(F_0)$. Thus, under H_0, the asymptotic levels are α as advertized. Indeed, Monte Carlo results in Table 1 and in Boos and Brownie (1986) show that the levels are close to α even in small samples. Moreover the results in Sections 2 and 3 imply that under H_a the power tends to one.

In this context we note that bootstrap resampling from the samples $X_1-\hat{\mu}_1,\ldots,X_m-\hat{\mu}_1$ and $Y_1-\hat{\mu}_2,\ldots,Y_n-\hat{\mu}_2$ separately is exactly the same as plan (II) since scale statistics are invariant to location shifts. A valid plan, however, is to bootstrap from

(IV) the samples $X_1/\hat{\theta}_1,\ldots,X_m/\hat{\theta}_1$ and $Y_1/\hat{\theta}_2,\ldots,Y_n/\hat{\theta}_2$ separately, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the scale estimators to be compared.

This plan will have the correct level asymptotically (even without the location-scale family assumption) and also appropriate power under alternatives.

The pooled procedure (III) we are suggesting, however, has an important advantage in small to moderate samples. Table 1 illustrates with Monte Carlo estimates of the true level of four tests based on the F statistic $\hat{\theta}_1/\hat{\theta}_2$. The four rows of the table differ only in how the critical values were chosen. Row 1 uses an F(m-1,n-1) critical value and is displayed to remind the reader of the nonrobustness of the classical procedure. Rows 2 and 3 pertain to the permutation and bootstrap resampling from $\{X_1-\bar{X},\ldots,X_m-\bar{X},\ Y_1-\bar{Y},\ldots,Y_n-\bar{Y}\}$ which are studied in Sections 2 and 3 of this paper. Row 4 is the separate rescaled plan (IV).

At the normal distribution all four tests in Table 1 hold their levels well. But at the Laplace distribution with density $f(x) = \frac{1}{2}\exp(-|x|)$ and at the Extreme Value distribution with distribution function $F(x) = \exp\{-\exp(-x)\}$, rows 2 and 3 are much better than rows 1 and 4. Apparently pooling is quite useful for these sample sizes and distributions. For this reason we have chosen to focus on the pooled procedures.

Table 1. Estimated levels of one-sided $\alpha = .05$ level tests based on the ratio of sample variances. Samples sizes: (m,n) = (10,15).

Method of obtaining critical values	Normal	Laplace	Extreme Value
F table	.046	.156	.099
Permutation, plan (III)	.048	.054	.063
Bootstrap, plan (III)	.050	.055	.059
Bootstrap, plan (IV)	.059	.122	.100

Note: Entries are based on 1000 Monte Carlo replications and have approximate standard deviation $((.95)(.05)/1000)^{1/2} = .007$. Permutation and bootstrap procedures are based on 500 resamples for each Monte Carlo replication.

From the above discussion it is clear that resampling from the location aligned and then combined sample is favorable. The basic results of this paper are to give the relevant asymptotic normal theory for resampling U- and V-statistics from this plan (III). We focus on U- and V-statistics because they include the sample variance which was the original motivating statistic and they often appear naturally in connection with other statistics. The discussion is restricted to kernels of degree two.

Results for bootstrapping are given in Section 2 and permutational results are in Section 3. The bootstrap results rely on techniques found in Bickel and Freedman (1981) and on some new theorems in the Appendix concerning almost sure convergence of U- and V-statistics with estimated parameters. The permutational results require additional theory along with standard permutational limit theorems for U-statistics (Puri and Sen (1971)). Section 4 gives further insight into the motivation and need for our approach and Section 5 has a numerical example. A short summary with possible extensions is given in Section 6.

We end this Introduction with a Proposition on T_{mn} and its associated V-statistics. For the two independent samples $X_1,\dots,X_m \sim F_1$ and $Y_1,\dots,Y_n \sim F_2$ with empirical distribution functions $\hat{F}_1(x) = m^{-1}\Sigma_{i=1}^{m} I(X_i \leq x)$ and $\hat{F}_2(x) = n^{-1}\Sigma_{i=1}^{n} I(Y_i \leq x)$, consider the standardized V-statistics $V_1 = m^{½}[\theta(\hat{F}_1) - \theta(F_1)]$ and $V_2 = n^{½}[\theta(\hat{F}_2) - \theta(F_2)]$, where $\theta(\cdot)$ is a simple von Mises functional of the form

$$(1.3) \qquad \theta(K) = \iint h(x,y)dK(x)dK(y)$$

for symmetric kernel h and any distribution function K. The limiting distribution of these standardized V-statistics can easily be derived via the differential approach (Serfling (1980)) by approximating $\theta(\hat{F}_i) - \theta(F_i)$ by $\theta(F_i;\hat{F}_i - F_i)$ where

$$\theta(F_1;\hat{F}_1 - F_1) = \frac{1}{m}\sum_{i=1}^{m} \psi(X_i,F_1) \text{ and } \theta(F_2;\hat{F}_2 - F_2) = \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i,F_2)$$

and

$$(1.4) \qquad \psi(x,K) = 2\left[\int h(x,y)dK(y) - \theta(K)\right] .$$

This approach leads directly to the asymptotic normality of (V_1,V_2) with asymptotic variances $\sigma^2(F_1)$ and $\sigma^2(F_2)$, respectively, where

$$(1.5) \qquad \sigma^2(K) = \int\psi^2(x,K)dK(x) .$$

From this joint asymptotic normality and assuming $h(x,y) > 0$, we can get the H_0: $\theta(F_1) = \theta(F_2)$ limiting distribution of T_{mn}. Formally, we have

<u>**PROPOSITION.**</u> Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ be iid with respective distribution functions F_1 and F_2. Assume that for $i = 1,2$

$$\iint h^2(x,y)dF_i(x)dF_i(y) < \infty , \quad \int h^2(x,x)dF_i(x) < \infty .$$

If $m,n \to \infty$ such that $m/(m+n) \to \lambda \in (0,1)$, then

$$(V_1,V_2) \xrightarrow{d} N(\underset{\sim}{0}, \operatorname{diag}(\sigma^2(F_1), \sigma^2(F_2)) .$$

Further, if $h(x,y) > 0$, then

$$T_{mn} - \left(\frac{mn}{m+n}\right)^{1/2} \log\frac{\theta(F_1)}{\theta(F_2)} \xrightarrow{d} N\left(0, \frac{(1-\lambda)\sigma^2(F_1)}{[\theta(F_1)]^2} + \frac{\lambda\sigma^2(F_2)}{[\theta(F_2)]^2}\right) .$$

2. LOCATION ALIGNED RESAMPLING WITH REPLACEMENT (BOOTSTRAP)

Consider the combined sample, aligned for location:

$$(Z_{N1},\ldots,Z_{NN}) = (X_1 - \hat{\mu}_1,\ldots,X_m - \hat{\mu}_1, Y_1 - \hat{\mu}_2,\ldots,Y_n - \hat{\mu}_2)$$

where $N = m + n$. Denote

$$H_{mn}(x) = N^{-1} \sum_{i=1}^{N} I(Z_{Ni} \le x) \quad .$$

Let $X^*_{N1},\ldots,X^*_{Nm}$, $Y^*_{N1},\ldots,Y^*_{Nn}$ be conditionally independent (given $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$) with common distribution function H_{mn}.

The idea is to compare the limit law of T_{mn} (see the Proposition in Section 1) with the conditional limit law of

$$T^*_{mn} = \left(\frac{mn}{N}\right)^{1/2} [\log \theta(\hat{F}^*_1) - \log \theta(\hat{F}^*_2)] \ ,$$

where $\hat{F}^*_1(x) = m^{-1} \Sigma^m_{i=1} I(X^*_{Ni} \le x)$ and $\hat{F}^*_2(x) = n^{-1} \Sigma^n_{i=1} I(Y^*_{Ni} \le x)$. As in the Proposition, it is easiest to work with the standardized V-statistics. Here we denote them by $V^*_1 = m^{1/2}[\theta(\hat{F}^*_1) - \theta(H_{mn})]$ and $V^*_2 = n^{1/2}[\theta(\hat{F}^*_2) - \theta(H_{mn})]$. Let P^*, E^*, Var^*,... denote probability, expectation, variance,... under H_{mn}.

The following set of assumptions will come into play in this section and in Section 3.

ASSUMPTIONS (A)

(A.0) $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ are iid F_1 and F_2, respectively.

(A.1) $m/N \to \lambda \in (0,1)$ as $m,n \to \infty$

(A.2) $\hat{\mu}_i \xrightarrow{a.s.} \mu_i$, $i = 1,2$.

For some $d > 0$ and $i = 1,2$:

(A.3) $\displaystyle\iint \sup_{|\gamma-\mu_i| \le d} h^2(x-\gamma,y-\gamma)dF_i(x)dF_i(y) < \infty$

(A.4) $\displaystyle\int \sup_{|\gamma-\mu_i| \le d} h^2(x-\gamma,x-\gamma)dF_i(x) < \infty$

(A.5) $$\iint \sup_{\gamma \in D(\mu,d)} h^{2\delta}(x-\gamma_1, y-\gamma_2)dF_1(x)dF_2(y) < \infty, \text{ for some } \delta > 1$$

where $\gamma = (\gamma_1, \gamma_2)$ and $D(\mu, d)$ is a sphere with radius d centered at $\mu = (\mu_1, \mu_2)$.

(A.6) $$\lim_{d\to 0} \iint \Big[\sup_{|\gamma-\mu_i| \le d} |h(x-\gamma, y-\gamma) - h(x-\mu_i, y-\mu_i)| \Big]^2 dF_i(x)dF_i(y) = 0$$

(A.7) $$\lim_{d\to 0} \int \Big[\sup_{|\gamma-\mu_i| \le d} |h(x-\gamma, x-\gamma) - h(x-\mu_i, x-\mu_i)| \Big]^2 dF_i(x) = 0$$

(A.8) $$\lim_{d\to 0} \iint \Big[\sup_{\gamma \in D(\mu,d)} |h(x-\gamma_1, y-\gamma_2) - h(x-\mu_1, y-\mu_2| \Big]^{2\delta} dF_1(x)dF_2(y) = 0,$$

for some $\delta > 1$.

Remarks. (1) The above conditions are natural in the context of U- and V-statistics with estimated parameters, see e.g., Randles (1982).

(2) Conditions (A.5) and (A.8) can be slightly modified in the sense of the moment condition in Theorem 1 of Sen (1977). However the present conditions are more attractive to check.

(3) If $\theta(\cdot)$ is the variance functional, then (A.3)-(A.8) hold if F_1 and F_2 have finite $4 + \varepsilon$ moments for some $\varepsilon > 0$. In that case, though, a direct proof shows that the following Theorem 1 is true with only the assumption of finite 4^{th} moments.

(4) The results of Theorem 1 are also true if (V_1^*, V_2^*) are replaced by the corresponding U-statistics.

THEOREM 1. Assume (A). Then almost surely (with respect to $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$)

$$(V_1^*, V_2^*) \xrightarrow{d} N(\underset{\sim}{0}, \operatorname{diag}(\sigma^2(H), \sigma^2(H))) ,$$

where $\sigma^2(H) = \int \psi^2(x,H)dH(x)$ and $H(x) = \lambda F_1(x + \mu_1) + (1 - \lambda)F_2(x + \mu_2)$. If

$h(x,y) > 0$, then almost surely

$$T^*_{mn} \xrightarrow{d} N(0,\ \sigma^2(H)/[\theta(H)]^2)\ .$$

PROOF OF THEOREM 1. Since the convergence of T^*_{mn} follows directly from that of (V_1^*,V_2^*), we only show the convergence of (V_1^*,V_2^*). To get the conditional limit law of $V_1^* = m^{1/2}[\theta(\hat{F}_1^*) - \theta(H_{mn})]$, we would show that

$$(2.1) \qquad m^{1/2}\left[\theta(\hat{F}_1^*) - \theta(H_{mn}) - \frac{1}{m}\sum_{i=1}^{m}\psi(X^*_{Ni},H_{mn})\right] = R_{mn} \xrightarrow{P^*} 0$$

$$(2.2) \qquad m^{-1/2}\sum_{i=1}^{m}\left[\psi(X^*_{Ni},H_{mn}) - \left\{\psi(X^*_{Ni},H) - \int\psi(x,H)dH_{mn}(x)\right\}\right] \xrightarrow{P^*} 0$$

$$(2.3) \qquad m^{-1/2}\sum_{i=1}^{m}\left[\psi(X^*_{Ni},H) - \int\psi(x,H)dH_{mn}(x)\right] \xrightarrow{d^*} N(0;\sigma^2(H))\ .$$

The joint convergence of (V_1^*,V_2^*) is only slightly more work. First, the analogues for (2.1) and (2.2) for V_2^* will follow the proof below for V_1^*. Call the left-hand side of (2.3) L_m and the analogous V_2^* quantity L_n. Then the joint convergence of (V_1^*,V_2^*) follows from the Cramér–Wold device if

$$c_1L_m + c_2L_n \xrightarrow{d^*} N(0,\ (c_1 + c_2)\sigma^2(H))$$

for every pair $(c_1,c_2) \in R^2$. But since L_m and L_n are conditionally independent, we need only show (2.3) and its V_2^* analogue.

PROOF of (2.1): As in Bickel and Freedman (1981, p. 1202) we have, for some constants C_1 and C_2

$$E^*(R^2_{mn}) \le \frac{C_1}{m}\,E^*\left[h^2(X^*_{N1},X^*_{N2})\right] + \frac{C_2}{m^2}\,E^*\left[h^2(X^*_{N1},X^*_{N1})\right]\ .$$

To show (2.1) it suffices to establish that $E^*(R^2_{mn}) \xrightarrow{a.s.} 0$. This is obtained by proving that

$$(2.4) \qquad E^*\left[h^2(X^*_{N1},X^*_{N2})\right] \xrightarrow{a.s.} \iint h^2(x,y)dH(x)dH(y)$$

and

$$(2.5) \qquad E^*\left[h^2(X^*_{N1},X^*_{N1})\right] \xrightarrow{a.s.} \int h^2(x,x)dH(x) \ .$$

To show (2.4), note that

$$E^*\left[h^2(X^*_{N1},X^*_{N2})\right] = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N} h^2(Z_{Ni},Z_{Nj})$$

$$= \left(\frac{m}{N}\right)^2 m^{-2}\sum_{i=1}^{m}\sum_{j=1}^{m} h^2(X_i - \hat{\mu}_1,\ X_j - \hat{\mu}_1)$$

$$+ \left(\frac{n}{N}\right)^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} h^2(Y_i - \hat{\mu}_2,\ Y_j - \hat{\mu}_2)$$

$$+ 2\left(\frac{mn}{N^2}\right)^2 (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} h^2(X_i - \hat{\mu}_1,\ Y_j - \hat{\mu}_2) \ .$$

The first two terms contain one-sample V-statistics with estimated parameters and the third term contains a two-sample V-statistic with estimated parameters. Their a.s. convergence can be established by using Lemma 1 in the Appendix. As an example we show that

$$(mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} h^2(X_i - \hat{\mu}_1,\ Y_j - \hat{\mu}_2) \xrightarrow{a.s.} \iint h^2(x-\mu_1,\ y-\mu_2)dF_1(x)dF_2(y) \ .$$

The conditions (i) and (ii) of Lemma 1 are satisfied by (A.2) and (A.5). Condition (iii) follows from the inequality

$$\iint \Big[\sup_{\gamma \in D(\mu,d)} |h^2(x-\gamma_1,\ y-\gamma_2) - h^2(x-\mu_1,\ y-\mu_2|\Big]^{\delta} dF_1(x)dF_2(y)$$

$$\leq 2^{\delta-1} \iint \Big[\sup_{\gamma \in D(\mu,d)} |h(x-\gamma_1,\ y-\gamma_2) - h(x-\mu_1,\ y-\mu_2)| \Big]^{\delta}$$

$$\cdot \Big[|\sup_{\gamma \in D(\mu,d)} h(x-\gamma_1,\ y-\gamma_2)|^{\delta} + |h(x-\mu_1,\ y-\mu_2)|^{\delta} \Big] dF_1(x) dF_2(y)$$

together with Hölder's inequality and (A.5) and (A.8).

Similarly we obtain (2.5).

PROOF of (2.2): To establish (2.2) it is easily seen that we only have to show the a.s. convergence to zero of

$$E^*\Big[(\psi(X^*_{N1},H_{mn}) - \psi(X^*_{N1},H))^2\Big]$$

$$= N^{-1} \sum_{i=1}^{N} \Big[\psi(Z_{Ni},H_{mn}) - \psi(Z_{Ni},H)\Big]^2$$

$$= \frac{m}{N}\frac{1}{m} \sum_{i=1}^{m} \Big[\psi(X_i - \hat{\mu}_1,H_{mn}) - \psi(X_i - \hat{\mu}_1,H)\Big]^2$$

$$+ \frac{n}{N}\frac{1}{n} \sum_{j=1}^{n} \Big[\psi(Y_j - \hat{\mu}_2,H_{mn}) - \psi(Y_j - \hat{\mu}_2,H)\Big]^2$$

$$\equiv \frac{m}{N} B_1 + \frac{n}{N} B_2 .$$

Under (A) it can be shown that $B_i \xrightarrow{a.s.} 0$, $i = 1,2$ (see Lemma 3 in the Appendix).

PROOF of (2.3): Using the Berry-Esseen theorem, it follows as in Singh (1981) that (2.3) holds if we show

(2.6) $$m^{-1/2} \Big\{E^*|\psi(X^*_{N1},H)|^3 + |E^*[\psi(X^*_{N1},H)]|^3\Big\} \xrightarrow{a.s.} 0$$

and

(2.7) $$\mathrm{Var}^*[\psi(X^*_{N1},H)] = E^*[\psi^2(X^*_{N1},H)] - (E^*[\psi(X^*_{N1},H)])^2 \xrightarrow{a.s.} \sigma^2(H) .$$

Under (A), Lemma 1 of the Appendix can be applied to get

$$E^*[\psi(X^*_{N1},H)] \xrightarrow{a.s.} \int\psi(x,H)dH(x) = 0$$

and

$$E^*[\psi^2(X^*_{N1},H)] \xrightarrow{a.s} \int\psi^2(x,H)dH(x) = \sigma^2(H) .$$

Hence (2.7) is valid. To get (2.6), it remains to show that $m^{-\frac{1}{2}}\ E^*|\psi(X^*_{N1},H)|^3 \xrightarrow{a.s.} 0$. Therefore note that

$$m^{-\frac{1}{2}}\ E^*|\psi(X^*_{N1},H)|^3$$

$$= \frac{m}{N}\, m^{-3/2} \sum_{i=1}^{m} |\psi(X_i - \hat{\mu}_1,H)|^3 + \left(\frac{n}{N}\right)^{3/2}\left(\frac{m}{N}\right)^{-\frac{1}{2}} n^{-3/2} \sum_{j=1}^{n} |\psi(Y_j - \hat{\mu}_2,H)|^3$$

$$= \frac{m}{N} D_1 + \left(\frac{n}{N}\right)^{3/2}\left(\frac{m}{N}\right)^{-\frac{1}{2}} D_2 .$$

Finally, $D_i \xrightarrow{a.s.} 0$, $i = 1,2$, by an application of Lemma 2 in the Appendix (Marcinkiewicz's strong law of large numbers for sums with estimated parameters), since it can be shown that the conditions of Lemma 2 are satisfied under (A).

3. LOCATION ALIGNED RESAMPLING WITHOUT REPLACEMENT (PERMUTATION)

As an alternative to the resampling scheme of Section 2, we now consider resampling without replacement from the combined sample, aligned for location. In this case $\tilde{X}_{N1},\ldots,\tilde{X}_{Nm}$ are sampled without replacement from $Z_{N1},\ldots,Z_{NN}$ and $\tilde{Y}_{N1},\ldots,\tilde{Y}_{Nn}$ denote the remaining Z_{Ni}'s after selecting $\tilde{X}_{N1},\ldots,\tilde{X}_{Nm}$. Our interest is now in the conditional limit law of

$$\tilde{T}_{mn} = \left(\frac{mn}{N}\right)^{\frac{1}{2}} [\log U'_{mn} - \log U''_{mn}] ,$$

where

$$U'_{mn} = \binom{m}{2}^{-1} \sum_{1 \le i < j \le m} \sum h(\tilde{X}_{Ni}, \tilde{X}_{Nj})$$

$$U''_{mn} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \sum h(\tilde{Y}_{Ni}, \tilde{Y}_{Nj}) \quad .$$

We will use the notation $\tilde{P}$, $\tilde{E}$, $\widetilde{Var}$,... for probability, expectation, variance,... under the above resampling scheme.

Note that if $h(x,y) = (x-y)^2/2$, U'_{mn} and U''_{mn} are the sample variances of the $\tilde{X}_{Ni}$ and $\tilde{Y}_{Ni}$, respectively, with divisors m-1 and n-1. Here we are giving a theorem about U-statistics because the notation is easier, but Theorem 2 holds for the related V-statistics as well. Note also that we only give the limiting distribution of U'_{mn} and $\tilde{T}_{mn}$ since U'_{mn} and U''_{mn} are functionally dependent.

The following notation will be used in the statement and proof of Theorem 2.

$$\theta_N = \binom{N}{2}^{-1} \sum_{1 \le i < j \le N} \sum h(Z_{Ni}, Z_{Nj})$$

$$\tilde{\theta}_N = N^{-1} \sum_{i=1}^{N} \int h(Z_{Ni}, y)\, dH(y)$$

$$\psi_n(Z_{Ni}) = 2\left[\frac{1}{N-1} \sum_{\substack{j=1 \\ j \ne i}}^{N} h(Z_{Ni}, Z_{Nj}) - \theta_N\right] , \quad i = 1, \ldots, N \quad .$$

THEOREM 2. Assume (A). Then

$$m^{1/2}(U'_{mn} - \theta_N) \xrightarrow{\tilde{d}} N(0,\ (1-\lambda)\sigma^2(H))$$

and, if $h(x,y) > 0$, then

$$\tilde{T}_{mn} \xrightarrow{\tilde{d}} N(0,\ \sigma^2(H)/[\theta(H)]^2) \quad ,$$

where $\sigma^2(H)$ is as in Theorem 1.

PROOF OF THEOREM 2. First, let R'_{mn} and R''_{mn} be defined by

$$m^{1/2}(U'_{mn} - \theta_N) = m^{-1/2} \sum_{i=1}^{m} \psi_N(\tilde{X}_{Ni}) + R'_{mn} ,$$

and

$$n^{-1/2} (U''_{mn} - \theta_N) = n^{-1/2} \sum_{j=1}^{n} \psi_N(\tilde{Y}_{Nj}) + R''_{mn} .$$

By adding and subtracting and Taylor expansion, we have

$$\left(\frac{N}{mn}\right)^{1/2} \tilde{T}_{mn} = \frac{1}{\theta_N} \left[\frac{1}{m} \sum_{i=1}^{m} \psi_N(\tilde{X}_{Ni}) - \frac{1}{n} \sum_{j=1}^{n} \psi_N(\tilde{Y}_{Nj})\right]$$

$$+ \left(U'_{mn} - \theta_N\right)\left(\frac{1}{\hat{\theta}_m} - \frac{1}{\theta_N}\right) - \left(U''_{mn} - \theta_N\right)\left(\frac{1}{\hat{\theta}_n} - \frac{1}{\theta_N}\right)$$

$$+ \frac{1}{\theta_N} \left(m^{-1/2} R'_{mn} - n^{-1/2} R''_{mn}\right) ,$$

where $\hat{\theta}_m$ lies between U'_{mn} and θ_N and $\hat{\theta}_n$ lies between U''_{mn} and θ_N. The key step here is to note that

$$(3.1) \qquad \sum_{i=1}^{N} \psi_N(Z_{Ni}) = \sum_{i=1}^{m} \psi_N(\tilde{X}_{Ni}) + \sum_{j=1}^{n} \psi_N(\tilde{Y}_{Nj}) = 0 .$$

Thus, substituting for $\Sigma_{j=1}^{n}\psi_N(\tilde{Y}_{Nj})$ we have

$$\tilde{T}_{mn} = \left(\frac{N}{n}\right)^{1/2} \frac{1}{\theta_N} \left[m^{-1/2} \sum_{i=1}^{m} \psi_N(\tilde{X}_{Ni})\right] + R_{mn} ,$$

where R_{mn} is the collection of remaining terms. The convergence of $m^{1/2}(U'_{mn} - \theta_N)$ follows if we show that

$$(3.2) \qquad R'_{mn} \xrightarrow{\tilde{P}} 0$$

$$(3.3) \qquad m^{-1/2} \sum_{i=1}^{m} \left[\psi_N(\tilde{X}_{Ni}) - 2\left\{ \int h(\tilde{X}_{Ni},y)dH(y) - \tilde{\theta}_N\right\}\right] \xrightarrow{\tilde{P}} 0$$

$$(3.4) \qquad 2m^{-\frac{1}{2}} \sum_{i=1}^{m} \left[\int h(\tilde{X}_{Ni},y)dH(y) - \tilde{\theta}_N \right] \xrightarrow{d} N(0;\ (1-\lambda)\sigma^2(H)) \quad .$$

The convergence of $n^{\frac{1}{2}}(U''_{mn} - \theta_N)$ follows similarly by showing that $R''_{mn} \xrightarrow{p} 0$ and that the appropriate analogues of (3.3) and (3.4) are valid. Using (A), an application of Lemma 1 in the Appendix shows that $\theta_N \xrightarrow{a.s.} \theta(H)$. (For a very similar argument, also see the end of the proof of Lemma 3 in the Appendix.) Hence $R_{mn} \xrightarrow{p} 0$. Finally since $\theta_N(n/N)^{\frac{1}{2}}\ \tilde{T}_{mn} - m^{\frac{1}{2}}(U'_{mn} - \theta_N) = R_{mn} - R'_{mn} \xrightarrow{p} 0$ and using again that $\theta_N \xrightarrow{a.s.} \theta(H)$, the result for $\tilde{T}_{mn}$ follows.

PROOF of (3.2): We only give the proof for R'_{mn} since R''_{mn} is treated analogously. Since $R'_{mn} \xrightarrow{p} 0$ follows by showing that $\tilde{E}(R'^2_{mn}) \xrightarrow{a.s.} 0$, the following inequality which can be obtained by straightforward but very tedious calculations, is basic. For some constant C

$$\tilde{E}(R'^2_{mn}) \le \frac{C}{m} \left[\frac{1}{N(N-1)} \sum_{1\le i\ne j\le N}\sum h^2(Z_{Ni},Z_{Nj})\right.$$

$$+ \frac{1}{N(N-1)(N-2)} \sum_{1\le i\ne j\ne k\le N}\sum\sum h(Z_{Ni},Z_{Nj})h(Z_{Ni},Z_{Nk})$$

$$\left. + \frac{1}{N(N-1)(N-2)(N-3)} \sum_{1\le i\ne j\ne k\ne \ell\le N}\sum\sum\sum h(Z_{Ni},Z_{Nj})h(Z_{Nk},Z_{N\ell})\right] \quad .$$

Now argue as in the proof of (2.1), using Lemma 1 in the Appendix, to show the a.s. convergence of each of the three terms within square brackets in the right-hand side of the above inequality.

PROOF of (3.3): With $\pi_N(\tilde{X}_{Ni}) = \psi_N(\tilde{X}_{Ni}) - 2\{\int h(\tilde{X}_{Ni},y)dH(y) - \tilde{\theta}_N\}$, we have that $\tilde{E}[\pi_N(\tilde{X}_{Ni})] = 0$ and that

$$\tilde{Var}\left[m^{-\frac{1}{2}} \sum_{i=1}^{m} \pi_N(\tilde{X}_{Ni})\right] = \left(1 - \frac{m}{N}\right) \tilde{Var}\left[\pi_N(\tilde{X}_{N1})\right]$$

$$\leq \widetilde{\mathrm{Var}}\left[\pi_N(\tilde{X}_{N1})\right] = \tilde{E}\left[\pi_N^2(\tilde{X}_{N1})\right] \quad .$$

Hence, we only have to show that $\tilde{E}[\pi_N^2(\tilde{X}_{N1})] \xrightarrow{a.s.} 0$. Therefore, note that

$$\tilde{E}\left[\pi_N^2(\tilde{X}_{N1})\right] = \frac{m}{N}\frac{1}{m}\sum_{i=1}^{m}\left[\psi_N(X_i - \hat{\mu}_1) - 2\{\int h(X_i - \hat{\mu}_1, y)dH(y) - \tilde{\theta}_N\}\right]^2$$

$$+ \frac{n}{N}\frac{1}{n}\sum_{j=1}^{n}\left[\psi_N(Y_j - \hat{\mu}_2) - 2\{\int h(Y_j - \hat{\mu}_2, y)dH(y) - \tilde{\theta}_N\}\right]^2$$

$$= \frac{m}{N}P_1 + \frac{n}{N}P_2 \; .$$

The result follows since, given the regularity conditions, it can be shown that $P_i \xrightarrow{a.s.} 0$, $i = 1,2$ (see Lemma 4 in the Appendix).

PROOF of (3.4): Writing

$$\frac{2}{m}\sum_{i=1}^{m}\left[\int h(\tilde{X}_{Ni}, y)dH(y) - \tilde{\theta}_N\right] = \sum_{i=1}^{N} b_{Ni}\zeta_{Ni}$$

it is clear that we have to deal with a linear permutation statistic as defined e.g. in Puri and Sen (1971, p. 73), where (using their notation)

(3.5) $\quad B_N = (b_{N1},\ldots,b_{Nm}, b_{Nm+1},\ldots,b_{NN}) = (\frac{2}{m},\ldots,\frac{2}{m}, 0,\ldots,0)$

(3.6) $\quad A_N = (A_{N1},\ldots,A_{NN})$ with $A_{Ni} = \int h(Z_{Ni}, y)dH(y) - \tilde{\theta}_N$.

Therefore (3.4) follows from Theorem 3.4.1 in Puri and Sen (1971) if we show that B_N satisfies their relation (3.4.6) (the Wald-Wolfowitz condition), that A_N satisfies their relation (3.4.8) (Hoeffding's form of the Noether condition) a.s. and that $\tilde{\sigma}_N^2$ as defined in their relation (3.4.21) tends to $(1 - \lambda)\sigma^2(H)$ a.s. By direct calculation we get for B_N given by (3.5) that $\mu_{\tau,N}(B_N)/[\mu_{2,N}(B_N)]^{\tau/2} = O(1)$, $\tau = 3, 4,\ldots$. To get (3.4.8) of Puri and Sen

(1971) for A_N defined in (3.6), we have to show

$$(3.7) \qquad \frac{N^{-1}\sum_{i=1}^{N} \left| \int h(Z_{Ni},y)dH(y) - \tilde{\theta}_N \right|^{2+\varepsilon}}{\left[N^{-1}\sum_{i=1}^{N}\left[\int h(Z_{Ni},y)dH(y) - \tilde{\theta}_N\right]^2\right]^{(2+\varepsilon)/2}} = o(N^{\varepsilon/2}) \quad \text{a.s.}$$

We first show that

$$(3.8) \qquad N^{-1}\sum_{i=1}^{N} \left|\int h(Z_{Ni},y)dH(y) - \tilde{\theta}_N\right|^{2+\varepsilon} = o(N^{\varepsilon/2}) \quad \text{a.s.}$$

We therefore rewrite the left-hand side of (3.8) as

$$\frac{m}{N}\frac{1}{m}\sum_{i=1}^{m} \left|\int h(X_i - \hat{\mu}_1,y)dH(y) - \tilde{\theta}_N\right|^{2+\varepsilon}$$

$$+ \frac{n}{N}\frac{1}{n}\sum_{j=1}^{n} \left|\int h(Y_j - \hat{\mu}_2,y)dH(y) - \tilde{\theta}_N\right|^{2+\varepsilon} .$$

Hence, using Minkowski's inequality, it follows that for (3.8) to hold, it is sufficient to show that

$$(3.9) \qquad m^{-\frac{2+\varepsilon}{2}}\sum_{i=1}^{m} \left|\int h(X_i - \hat{\mu}_1,y)dH(y)\right|^{2+\varepsilon} \xrightarrow{\text{a.s.}} 0 \quad ,$$

$$(3.10) \qquad n^{-\frac{2+\varepsilon}{2}}\sum_{j=1}^{n} \left|\int h(Y_j - \hat{\mu}_2,y)dH(y)\right|^{2+\varepsilon} \xrightarrow{\text{a.s.}} 0 \quad ,$$

and to note that, under the assumed regularity conditions,

$$\tilde{\theta}_N = N^{-1}\sum_{j=1}^{N} \int h(Z_{Ni},y)dH(y) \xrightarrow{\text{a.s.}} \iint h(x,y)dH(x)dH(y)$$

by an application of Lemma 1 in the Appendix. To get (3.9) and (3.10), apply Lemma 2 in the Appendix with $\tau = 2/(2+\varepsilon)$. Using similar arguments it is easy to show that the denominator in (3.7) is a.s. convergent. Hence the order bound noted in (3.7) is valid. From Theorem 3.4.1 in Puri and Sen

(1971) it follows that the variance of $2m^{-1} \sum_{i=1}^{m} [\int h(\tilde{X}_{Ni}, y) dH(y) - \tilde{\theta}_N]$ is given by

$$\frac{mN^2}{N-1} \mu_{2,N}(B_N) \frac{1}{N} \sum_{i=1}^{N} A_{Ni}^2 = \frac{4n}{N-1} \left[\frac{m}{N} \frac{1}{m} \sum_{i=1}^{m} A_{Ni}^2 + \frac{n}{N} \frac{1}{n} \sum_{j=m+1}^{N} A_{nj}^2\right]$$

which converges a.s. to $(1-\lambda)\sigma^2(H)$.

4. COMPARISON OF ASYMPTOTIC VARIANCES

The Proposition in the Introduction shows that T_{mn} converges under H_0 to a normal distribution with variance $\sigma^2(F_1,F_2)/[\theta(F_1)]^2$, where

$$\sigma^2(F_1,F_2) = (1 - \lambda)\sigma^2(F_1) + \lambda\sigma^2(F_2) \quad . \tag{4.1}$$

Under the bootstrap and permutation resampling schemes, T^*_{mn} and $\tilde{T}_{mn}$ each converge to a normal distribution with variance $\sigma^2(H)/[\theta(H)]^2$, where $H(x) = \lambda F_1(x + \mu_1) + (1 - \lambda)F_2(x + \mu_2)$, and $\sigma^2(\cdot)$ is given by (1.5). We now illustrate two cases where these asymptotic variances are equal under H_0 and therefore justify the use of the resampling schemes.

<u>Case 1</u>. The variance kernel, $h(x,y) = (x - y)^2/2$. If μ_1 and μ_2 are the means of F_1 and F_2, respectively, and H_0 holds, i.e., variance of $F_1 = \theta(F_1) = \theta(F_2) =$ variance of F_2, then

$$\sigma^2(F_1,F_2)/[\theta(F_1)]^2 = (1 - \lambda)[\beta_2(F_1) - 1] + \lambda[\beta_2(F_2) - 1] \quad . \tag{4.2}$$

where $\beta_2(\cdot)$ is the kurtosis functional (1.2). Also, we have $\theta(H) = \lambda\theta(F_1) + (1 - \lambda)\theta(F_2) = \theta(F_1)$ and

$$\sigma^2(H)/[\theta(H)]^2 = \lambda[\beta_2(F_1) - 1] + (1 - \lambda)[\beta_2(F_2) - 1] \quad . \tag{4.3}$$

Thus, (4.2) and (4.3) are equal iff $\lambda = ½$ or $\beta_2(F_1) = \beta_2(F_2)$. Apparently equal sample sizes ($\lambda = ½$) give extra robustness against unequal kurtosis.

Case 2. Functionals having kernels of the form $h(x,y) = q(|x - y|)$. The most well-known examples are Gini's mean difference, $h(x,y) = |x - y|$, and obvious extensions to $|x - y|^p$, $p > 0$. In the location-scale model $F_i(x) = F_0((x-\mu_i)/\sigma_i)$ we have, under H_0: $\sigma_1 = \sigma_2$, that $\sigma^2(F_1) = \sigma^2(F_2)$ and that (4.1) is just $\sigma^2(F_1)$. Also, $\theta(H) = \theta(F_1)$ and $\sigma^2(H) = \sigma^2(F_1)$ so that $\sigma^2(F_1,F_2)/[\theta(F_1)]^2 = \sigma^2(H)/[\theta(H)]^2$.

5. EXAMPLE

A new drug regimen (B) was given to 16 subjects and a second independent group of 13 subjects received the standard drug regimen (A). At the end of one week each subject's status was assessed and a baseline measurement subtracted off yielding the raw data A: 0,8,1,-1,-3,-7,6,-1,14,0,-1, -1,-3, B: 6,-1,12,21,2,0,16,10,-1,-1,10,4,-22,-11,2,4. Primary interest was in location differences but a secondary concern might be with differences in variability. The sample variances are 29.7 and 104.2 and the usual F = 3.50 has a two-sided p value of .034 from an F table with 15 and 12 degrees of freedom. Since normality is suspect, we proceed to get an estimated p value from resampling. The means .92 and 3.19 were subtracted off and the set (-0.92,7.08,...,-1.19,0.81) used for resampling. For the bootstrap, 10,000 pairs of samples of $n_1 = 13$ and $n_2 = 16$ random variables were drawn with replacement and $\hat{p} = \{\#$ of sets with $|\log \theta(\hat{F}_1^*)/\theta(\hat{F}_2^*)| \geq \log 3.50]/10,000 = 0.13$, where $\theta(\cdot)$ is the variance functional. For the permutation approach, 10,000 pairs of samples were drawn where the first sample of size $n_1 = 13$ is drawn without replacement and the second sample consists of the remaining 16 members of the set. Here the two-sided $\hat{p} = 0.14$.

Another spread functional covered by Theorems 1 and 2 is Gini's mean difference $\theta(F) = \int\int |x_1 - x_2| dF(x_1)dF(x_2)$. Assumption (A) of Section 2 is satisfied for any distribution with $r > 2$ finite moments. The estimated bootstrap and permutation two-sided p values based on 10,000 resamples were $\hat{p} = .11$ and $\hat{p} = .12$ for this functional.

6. SUMMARY AND EXTENSIONS

We have provided asymptotic normal theory for resampling U- and V-statistics from aligned and pooled two-sample data. A key motivation was to provide consistent critical value estimates for tests of scale homogeneity. The theorems we have proved, however, cover only a small fraction of possible applications and here we would like to mention a few others for which the theory can be extended.

1) Many types of functionals can be used for the scale problem. These include classical L-, M-, and R-functionals such as the interquartile range. Each type functional will entail different technical details but the same general approach.

2) Two-sample test statistics for scale need not be of the form $\log[\theta(\hat{F}_1)/\theta(\hat{F}_2)]$. For example, consider Lehmann's (1951) U-statistic

$$\binom{m}{2}^{-1}\binom{n}{2}^{-1} \Sigma\Sigma I(|X_i - X_j| > |Y_i - Y_j|)$$

or Mood's rank test for scale based on aligned samples. Neither is asymptotically distribution-free and resampling could be useful. Note that our Theorems 1 and 2 do not cover two-sample U-Statistics but Lemma 1 in the Appendix does. See Boos (1986) for a recent discussion of Mood's test and a method of proof which can be extended to resampling.

3) Location-scale alignment is of interest when testing equality of skewness or kurtosis parameters. For example one could resample from

$$\left\{\frac{X_1-\hat{\mu}_1}{\hat{\sigma}_1}, \ldots, \frac{X_m-\hat{\mu}_1}{\hat{\sigma}_1}, \frac{Y_1-\hat{\mu}_2}{\hat{\sigma}_2}, \ldots, \frac{Y_n-\hat{\mu}_2}{\hat{\sigma}_2}\right\}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are estimators for location and scale of F_i, $i = 1, 2$. It is easy to see how Assumption (A) of Section 2 would need to be modified in order to get a theorem for U-statistics in this situation.

4) The basic theory can be extended to k-sample statistics such as Bartlett's test for homogeneity of scale. The details and conditions, though, can get quite cumbersome.

APPENDIX

Some results on the almost sure behavior of sums, having terms with an estimated parameter are given in the first part of this Appendix. Lemma 1 generalizes the strong law of large numbers for one-sample U-statistics with estimated parameters (Theorem 2.11 of Iverson and Randles, (1983)) to the k-sample situation. Lemma 2 extends the classical Marcinkiewicz strong law of large numbers (see e.g. Theorem 3.2.3 in Stout (1974)) to the case where the terms in the sum depend on an estimated parameter. A second part of the Appendix (Lemmas 3 and 4) deals with the details of the almost sure convergence to zero of B_i and P_i, $i = 1,2$, as defined in Sections 2 and 3.

Let $X_{11},\ldots,X_{1n_1};\ldots;X_{k1},\ldots,X_{kn_k}$ denote k independent samples from $F_1,\ldots,F_k$ and let, with $m_j \leq n_j$, $j = 1,\ldots,k$

$$\theta_{\underline{F}}(\gamma) = E_{\underline{F}}[h(X_{11},\ldots,X_{1m_1};\ldots;X_{k1},\ldots,X_{km_k};\gamma)] \quad ,$$

where h is a kernel depending on a (possibly multivariate) parameter γ and $\underline{F} = (F_1,\ldots,F_k)$. With $\underline{n} = (n_1,\ldots,n_k)$, the corresponding U-statistic is

$$U_{\underline{n}}(\gamma) = \left[\prod_{j=1}^{k} \binom{n_j}{m_j}\right]^{-1} \sum_{(1)}\cdots\sum_{(k)} h(X_{1i_{11}},\ldots,X_{1i_{1m_1}};\ldots;X_{ki_{k1}},\ldots,X_{ki_{km_k}};\gamma) \quad ,$$

where $\Sigma_{(j)}$ extends over all $\binom{n_j}{m_j}$ ordered m_j-tuples $1 \leq i_{j1} < \ldots < i_{jm_j} \leq n_j$ taken from $\{1,\ldots,n_j\}$, $j = 1,\ldots,k$.

The following lemma gives sufficient conditions for $U_n(\hat{\lambda})$ to converge to $\theta_{\underline{F}}(\lambda)$, where $\hat{\lambda}$ is an estimator for the true parameter λ.

LEMMA 1. Assume

(i) $\hat{\lambda} \xrightarrow{a.s.} \lambda$ as $\min(n_1,\ldots,n_k) \to \infty$

and for some $\delta > 1$:

(ii) $E_{\underline{F}}|h(X_{11},\ldots,X_{1m_1};\ldots;X_{k1},\ldots,X_{km_k};\lambda)|^{\delta} < \infty$

(iii) for some sphere $D(\lambda,d)$ with radius d, centered at λ,

$$\lim_{d\to 0} E_{\underline{F}}\Big[\sup_{\gamma\,\varepsilon\,D(\lambda,d)} |h(X_{11},\ldots,X_{1m_1};\ldots;X_{k1},\ldots,X_{km_k};\gamma)$$

$$- h(X_{11},\ldots,X_{1m_1};\ldots;X_{k1},\ldots,X_{km_k};\lambda|^{\delta}\Big] = 0$$

Then, as $\min(n_1,\ldots,n_k) \to \infty$,

$$U_n(\hat{\lambda}) \xrightarrow{a.s.} \theta_{\underline{F}}(\lambda) .$$

REMARKS. (1) The proof is similar to that of one-sample U-statistics with estimated parameters as given in Iverson (1982). In that case, δ can be taken to be 1. Taking $\delta > 1$ is sufficient to meet the condition of Theorem 1 in Sen (1977). Therefore, slightly weaker versions of (ii) and (iii) are possible. We could, for example, replace $E|h|^{\delta} < \infty$ in (ii) by $E[h(\log^{+}|h|)^{C-1}] < \infty$, etc.

(2) The conclusion of Lemma 1 also holds if we estimate $\theta_{\underline{F}}(\lambda)$ by the corresponding V-statistic, provided we allow in (ii) and (iii) that the m_j arguments in the j^{th} block of h are not necessarily distinct, j = 1,...k.

Our next lemma gives an extension of the Marcinkiewicz strong law of large numbers for sums of the form $\Sigma_{i=1}^{n} g(X_i;\hat{\lambda})$, where $X_1,\ldots,X_n$ are independent and identically distributed and $\hat{\lambda}$ is an estimator for some parameter λ.

LEMMA 2. Assume

(i) $\hat{\lambda} \xrightarrow{a.s.} \lambda$ as $n \to \infty$

and for some $0 < \tau < 1$

(ii) $E|g(X_1;\lambda)|^{\tau} < \infty$

(iii) for some sphere $D(\lambda,d)$ with radius d, centered at λ,

$$\lim_{d \to 0} E\left[\left[\sup_{\gamma \in D(\lambda,d)} |g(X_1;\gamma) - g(X_1;\lambda)|\right]^{\tau}\right] < \infty \quad .$$

Then, as $n \to \infty$,

$$n^{-1/\tau} \sum_{i=1}^{n} |g(X_i;\hat{\lambda})| \xrightarrow{a.s.} 0 \quad .$$

PROOF. It is sufficient to show that $Q_n(\hat{\lambda} - \lambda) \xrightarrow{a.s.} 0$, where

$$Q_n(s) = n^{-1/\tau} \sum_{i=1}^{n} \left[g(X_i;\lambda + s) - g(X_i;\lambda)\right] \quad .$$

For C a sphere centered at zero and with radius less than d, we have

$$P(|Q_n(\hat{\lambda} - \lambda)| > \varepsilon \text{ i.o})$$

$$\leq P(\sup_{s \in C} |Q_n(s)| > \varepsilon \text{ i.o}) + P(|\hat{\lambda} - \lambda| \notin C \text{ i.o}) \quad .$$

The second term is 0 since $\hat{\lambda}$ is a strongly consistent estimator for λ. Since

$$\sup_{s \in C} |Q_n(s)| \leq n^{-1/\tau} \sum_{i=1}^{n} \sup_{s \in C} |g(X_i;\lambda + s) - g(X_i;\lambda)|$$

and since (iii) allows us to apply the classical Marcinkiewicz strong law, we get $Q_n(\hat{\lambda} - \lambda) \xrightarrow{a.s.} 0$ as $n \to \infty$.

We conclude this Appendix with two lemmas on the quantities B_1, B_2 in Section 2 and P_1, P_2 in Section 3.

LEMMA 3. Under (A): $B_i \xrightarrow{a.s.} 0$, $i = 1,2$.

PROOF. We only show that $B_1 \xrightarrow{a.s.} 0$ since the proof for B_2 is similar. We have

$$B_1 = \frac{1}{m}\sum_{i=1}^{m}\Big\{2\Big[\int h(X_i-\hat{\mu}_1,y)dH_{mn}(y) - \theta(H_{mn})\Big]$$

$$- 2\Big[\int h(X_i-\hat{\mu}_1,y)dH(y) - \theta(H)\Big]\Big\}^2$$

$$= \frac{4}{m}\sum_{i=1}^{m}\Big\{\frac{m}{N}\frac{1}{m}\sum_{j=1}^{m} h(X_i-\hat{\mu}_1,X_j-\hat{\mu}_1) + \frac{n}{N}\frac{1}{n}\sum_{j=1}^{n} h(X_i-\hat{\mu}_1,Y_j-\hat{\mu}_2)$$

$$- \theta(H_{mn}) - \int h(X_i-\hat{\mu}_1,y)dH(y) + \theta(H)\Big\}^2$$

$$= 4\Big(\frac{m}{N}\Big)^2\Big[\frac{1}{m^3}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} h(X_i-\hat{\mu}_1,X_j-\hat{\mu}_1)h(X_i-\hat{\mu}_1,X_k-\hat{\mu}_1)\Big]$$

$$+ 4\Big(\frac{m}{N}\Big)^2\Big[\frac{1}{mn^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} h(X_i-\hat{\mu}_1,Y_j-\hat{\mu}_2)h(X_i-\hat{\mu}_1,Y_k-\hat{\mu}_2)\Big]$$

$$+ 8\Big(\frac{m}{N}\Big)\Big(\frac{n}{N}\Big)\Big[\frac{1}{m^2n}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{n} h(X_i-\hat{\mu}_1,X_j-\hat{\mu}_1)h(X_i-\hat{\mu}_1,Y_j-\hat{\mu}_2)\Big]$$

$$- 8\Big(\frac{m}{N}\Big)\Big[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} h(X_i-\hat{\mu}_1,X_j-\hat{\mu}_1)\int h(X_i-\hat{\mu}_1,y)dH(y)\Big]$$

$$- 8\Big(\frac{n}{N}\Big)\Big[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} h(X_i-\hat{\mu}_1,Y_j-\hat{\mu}_2)\int h(X_i-\hat{\mu}_1,y)dH(y)\Big]$$

$$+ 4\Big[\frac{1}{m}\sum_{i=1}^{m}\Big(\int h(X_i-\hat{\mu}_1,y)dH(y)\Big)^2\Big]$$

$$- 8\Big(\frac{m}{N}\Big)\Big(\theta(H_{mn}) - \theta(H)\Big)\Big[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} h(X_i-\hat{\mu}_1,X_j-\hat{\mu}_1)\Big]$$

$$- 8\left(\frac{n}{N}\right)\left(\theta(H_{mn}) - \theta(H)\right)\left[\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} h(X_i-\hat{\mu}_1, Y_j-\hat{\mu}_2)\right]$$

$$+ 8\left(\theta(H_{mn}) - \theta(H)\right)\left[\frac{1}{m} \sum_{i=1}^{m} \int h(X_i-\hat{\mu}_1, y)dH(y)\right]$$

$$+ 4\left(\theta(H_{mn}) - \theta(H)\right)^2 .$$

For each of the one- or two-sample V-statistics in square brackets one can verify that they converge almost surely using Lemma 1 and Condition (A). The sum of the first six terms converges almost surely to

$$4\lambda^2 E[h(X_1-\mu_1, X_2-\mu_1)h(X_1-\mu_1, X_3-\mu_1)]$$

$$+ 4(1 - \lambda)^2 E[h(X_1-\mu_1, Y_1-\mu_2)h(X_1-\mu_1, Y_2-\mu_2)]$$

$$+ 8\lambda(1 - \lambda)E[h(X_1-\mu_1, X_2-\mu_1)h(X_1-\mu_1, Y_1-\mu_2)]$$

$$- 8\lambda E[h(X_1-\mu_1, X_2-\mu_1)\int h(X_1-\mu_1, y)dH(y)]$$

$$- 8(1 - \lambda)E[h(X_1-\mu_1, Y_1-\mu_2)\int h(X_1-\mu_1, y)dH(y)]$$

$$+ 4E[(\int h(X_1-\mu_1, y)dH(y))^2]$$

and a straightforward calculation, using the definition of H, shows that this is equal to zero. Each of the four remaining terms tends to zero almost surely since the factor $\theta(H_{mn}) - \theta(H) \xrightarrow{a.s.} 0$; indeed

$$\theta(H_{mn}) = N^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} h(Z_{Ni}, Z_{Nj}) = \left(\frac{m}{N}\right)^2\left[m^{-2} \sum_{i=1}^{m} \sum_{j=1}^{m} h(X_i-\hat{\mu}_1, X_j-\hat{\mu}_1)\right]$$

$$+ \left(\frac{n}{N}\right)^2\left[n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(Y_i-\hat{\mu}_2, Y_j-\hat{\mu}_2)\right]$$

$$+ 2\left(\frac{m}{N}\right)\left(\frac{n}{N}\right)\left[m^{-1}n^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} h(X_i-\hat{\mu}_1, Y_j-\hat{\mu}_2)\right]$$

$$\xrightarrow{a.s.} \lambda^2 E[h(X_1-\mu_1, X_2-\mu_1)] + (1-\lambda)^2 E[h(Y_1-\mu_2, Y_2-\mu_2)]$$

$$+ 2\lambda(1-\lambda)E[h(X_1-\mu_1, Y_1-\mu_2)] = \theta(H) \ .$$

LEMMA 4. Under (A): $P_i \xrightarrow{a.s.} 0, \quad i = 1,2$.

PROOF. We can restrict to P_1 since the proof for P_2 is similar. We have

$$P_1 = \frac{1}{m} \sum_{i=1}^{m} \left\{\psi_N(X_i-\hat{\mu}_1) - 2\left[\int h(X_i-\hat{\mu}_1, y)\,dH(y) - \tilde{\theta}_N\right]\right\}^2$$

$$= \frac{4}{m} \sum_{i=1}^{m} \left\{\frac{m-1}{N-1} \frac{1}{m-1} \sum_{\substack{j=1 \\ j\neq i}}^{m} h(X_i-\hat{\mu}_1, X_j-\hat{\mu}_1)\right.$$

$$\left. + \frac{n}{N-1} \frac{1}{n} \sum_{j=1}^{n} h(X_i-\hat{\mu}_1, Y_j-\hat{\mu}_2) - \int h(X_i-\hat{\mu}_1, y)\,dH(y) - (\theta_N - \tilde{\theta}_N)\right\}^2 \ .$$

From here we proceed similarly as in Lemma 3.

REFERENCES

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1217.

Boos, D. D. (1986). Comparing k populations with linear rank statistics. *J. Amer. Statist. Assoc.* 81, 1018-1025.

Boos, D. D. and Brownie, C. (1986). Bootstrap methods for testing homogeneity of variances. Institute of Statistics Mimeo Series No. 1969, North Carolina State University, Raleigh.

Box. G. E. P. (1953). Non-normality and tests on variances. *Biometrika* 40, 318-335.

Iverson, H. K. (1982). Asymptotic properties of U-statistics with estimated parameters. Unpublished Ph.D. Thesis.

Iverson. H. K. and Randles, R. H. (1983). Convergence results for statistics with estimated paramters. Technical Report No. 213, Dept. of Statist., Univ. of Florida, Gainesville.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statist.* 22, 165-179.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* J. Wiley, New York.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* 10, 462-474.

Sen, P. K. (1977). Almost sure convergence of generalized U-Statistics. *Ann. Probability* 5, 287-290.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* J. Wiley, New York.

Singh, K. (1981). On the accuracy of Efron's bootstrap. *Ann. Statist.* 9, 1187-1195.

Stout, W. F. (1974). *Almost Sure Convergence.* Academic Press, New York.