The Annals of Statistics
1985, Vol. 13, No. 1, 116–132

EDGEWORTH CORRECTED PIVOTAL STATISTICS AND THE BOOTSTRAP

BY LAVY ABRAMOVITCH AND KESAR SINGH[1]

Rutgers University

A general procedure for multistage modification of pivotal statistics is developed to improve the normal approximation. Bootstrapping a first stage modified statistic is shown to be equivalent, in terms of asymptotic order, to the normal approximation of a second stage modification. Explicit formulae are given for some basic cases involving independent random samples and samples drawn without replacement. The Hodges-Lehmann deficiency is calculated to compare the regular t-statistic with its one-step correction.

1. Introduction and basic ideas. Let $\hat{\theta}_n$ be an estimator of a parameter θ_F, based on a random sample of size n from the population F. Consider the pivotal statistic $T = (\hat{\theta}_n - \theta_F)/v_n$ where v_n is the estimated standard error of $\hat{\theta}_n$. In order to use a statistic like T to form confidence intervals for θ, one requires the sampling distribution of T. Typically T approaches the standard normal distribution in law as n increases. In many interesting cases, the probability distribution of T also admits a valid expansion of the following form:

$$P(T \le x) = \Phi(x) + \textstyle\sum_{i=1}^{k} n^{-i/2} p_i(F, x)\phi(x) + o(n^{-k/2}) \tag{1}$$

where Φ and ϕ are the d.f. and the density of the standard normal distribution; the P_i's are certain polynomials in x whose coefficients depend upon the first few central moments of F. The number of terms in the expansion depend upon the number of finite absolute moments of the population F. The remainder term is typically $O(n^{-(k+1)2})$. Evidently, the knowledge of certain moments of F is required to use such an expansion for approximating $P(T \le x)$. Needless to say, an experimenter generally does not have this information unless a large scale survey has been conducted on the population on a previous occasion. The subject of this paper is the direct and indirect use of Edgeworth expansions to increase the asymptotic accuracy in situations where the population moments are unknown.

If the population moments are unknown, a natural approach is to substitute the appropriate sample moments in the r.h.s. of the above expansion. In this situation, no matter how small the remainder is in the original expansion, the error term in the resulting expansion after sample substitution is at least $O_p(n^{-1})$, unless $P_1(F, x) \equiv 0$, in which case it may be smaller. The same asymptotic order can typically be achieved by bootstrapping T. To be more specific, consider drawing repeated random samples of size n from the empirical population based on the original sample and recomputing T for each second stage sample, using

Received July 1983; revised September 1984.

[1] Research supported by NSF Grant MCS81-02341, 83-01082.

AMS 1980 *subject classifications.* Primary 62E20, 62G10, 62G15, 62G20.

Key words and phrases. Pivotal statistics, confidence intervals, hypothesis testing, Edgeworth expansions, bootstrap procedure, random sampling without replacement.

the original $\hat{\theta}_n$ in place of θ in each case. The histogram of a large number of these bootstrap values of T generally approximates the true distribution of T with a remainder of order $O_p(n^{-1})$ (see Efron, 1979, Singh, 1981, and Babu and Singh, 1983, for some background). A third approach to using the expansions constitutes the main topic of this paper. In this case one tries to modify T so that the modification does not contain terms of certain higher orders in its Edgeworth expansion. One interesting feature of this approach is that it can be used to reduce the remainder beyond $O(n^{-1})$.

While this manuscript was under preparation the authors came across two recent publications Hall (1983) and Withers (1983) which are discussed briefly later in this section.

Part of our approach used in modifying statistics like T is found in Johnson (1978). Johnson looked at Student's t-statistic specifically and without formal justification used a Cornish-Fisher expansion to modify the t-statistic so as to eliminate the effect of population skewness. Our general procedure, though different from Johnson's approach, gives essentially the same result when applied to the first step correction of Student's t-statistic. In addition, our general procedure immediately suggests higher order corrections and makes the idea of these modifications clear mathematically. A connection between these modifications and the bootstrap is also discussed. The basic modification of T is the topic of Theorem 1.

THEOREM 1. *Suppose that T admits an Edgeworth expansion*

$$P(T \leq x) = \Phi(x) + n^{-1/2}p(F, x) + o(n^{-1/2})$$

uniformly in x, where $p(F, x)$ is a polynomial in x whose coefficients possibly depend upon F. Let $\hat{p}_n$ be an estimator of $p(F, T)$ which satisfies the condition that

$$\text{(2)} \qquad \textit{for all } \varepsilon > 0, \quad P(|\hat{p}_n - p(F, T)| > \varepsilon) = o(n^{-1/2})$$

as $n \to \infty$. Then T_1 defined as

$$T_1 = T + n^{-1/2}\hat{p}_n$$

has the following Edgeworth expansion:

$$\text{(3)} \qquad P(T_1 \leq x) = \Phi(x) + o(n^{-1/2})$$

uniformly in x.

REMARK 1. If $p(F, x)$ is a polynomial in x whose coefficients depend on F only through its first r moments, if these coefficients as a function of these moments have bounded partial derivatives in a neighborhood of the true moments of F, and $\int |X|^{1.5r+\delta}\, dF < \infty$ for any $\delta > 0$, then

$$P(|p(F_n, T) - p(F, T)| > \varepsilon) = o(n^{-1/2})$$

for any $\varepsilon > 0$. Here F_n denotes the empirical c.d.f. Thus, in this case, one can take $T_1 = T + n^{-1/2}p(F_n, T)$.

All the proofs are deferred. The remainders in (2) and (3) are generally $O(1/n)$. In fact the Edgeworth expansion of T_1, if it exists, typically looks like

$$\Phi(x) + n^{-1}q(F, x)\phi(x) + o(n^{-1}).$$

And Theorem 1 suggests the following second stage corrections:

$$T_2 = T_1 + n^{-1}\hat{q}_n,$$

where $\hat{q}_n$ is an estimator of $q(F, T_1)$ satisfying

$$P(|\hat{q}_n - g(F, T_1)| > \varepsilon) = o(n^{-1})$$

for any $\varepsilon > 0$. Here too, $\hat{q}_n$ can be taken as $q(F_n, T_1)$ under certain restrictions. T_2 will not have a $1/n$ term in its own Edgeworth expansion. Theoretically this process can be carried out up to k-terms. However, with moderate sample practicality in mind it is probably not worth going beyond a two-step correction. The reason for this is that higher order corrections involve estimation of higher order moments. As is generally known, these are hard to estimate with reasonable accuracy. In Section 2, certain variations of T_1 and T_2 are discussed which may improve performance in the case of moderate sample size.

Is there any gain to be made in bootstrapping T_1 or T_2 instead of using the standard normal approximation? The answer turns out to be an encouraging "yes". The standard normal approximation to the distribution of T_1 is valid with a remainder of order $o(n^{-1/2})$, whereas the bootstrap distribution of T_1 approximates its true distribution with an error term $o(n^{-1})$; thus the bootstrap gives an extra step of accuracy. This fact is due to the bootstrap distribution of T_1 having the Edgeworth expansion

$$\Phi(x) + n^{-1}q(F_n, x)\phi(x) + o(n^{-1}) = \Phi(x) + n^{-1}q(F, x)\phi(x) + o(n^{-1}) \quad \text{(a.s.)}.$$

The validity of this expansion for the bootstrap distribution usually does not require additional assumptions because all the requirements on F_n are guaranteed for all large n by the same requirements on F. Thus bootstrapping T_1 basically amounts to a second-stage correction. This is an important fact, since it is usually very tedious to get hold of the exact expression for T_2. It is comforting to know that the asymptotic accuracy achieved by the normal approximation of T_2 is also obtained by a computer-based alternative, namely, bootstrapping T_1. Thus bootstrapping amounts to substitution of certain tedious analytical computations by the computer's brute force! Indeed, bootstrapping T_2 amounts to a third-stage correction.

The methods in this paper can be used to construct confidence intervals for θ_F which keep the same kind of asymptotic precision as discussed above for confidence levels. Let $-z_{\alpha/2} = \Phi^{-1}(\alpha/2)$ for some $\alpha \in (0, \frac{1}{2})$, so that $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. Let us confine ourselves to the case $T_1 = T + n^{-1/2}p(F_n, T)$. Define

$$a_1 = z_{\alpha/2} - n^{-1/2}p(F_n, z_{\alpha/2}),$$

and

$$a_2 = -z_{\alpha/2} - n^{-1/2}p(F_n, -z_{\alpha/2}).$$

Then the following result holds:

THEOREM 2. *If T admits a one-term Edgeworth expansion and the conditions of Remark 1 hold, then*

$$P(\theta_F < \hat{\theta}_n - a_1 v_n) = P(T_1 > z_{\alpha/2}) + o(n^{-1/2}) = \alpha/2 + o(n^{-1/2}) \tag{4}$$

and

$$P(\theta_F > \hat{\theta}_n - a_2 v_n) = P(T_1 < -z_{\alpha/2}) + o(n^{-1/2}) = \alpha/2 + o(n^{-1/2}). \tag{5}$$

Thus $[\hat{\theta}_n - a_1 v_n, \hat{\theta}_n - a_2 v_n]$ is a one-step corrected C.I. which leaves out $\alpha/2 + o(n^{-1/2})$ probability in each tail. Usually, $p(F, \cdot)$ is an even polynomial; thus $p(F_n, z_{\alpha/2}) = p(F_n, -z_{\alpha/2})$. In this case, the one-step modified interval has the same length as the naive C.I. $[\theta - zv_n, \theta + zv_n]$ based on the CLT. Note that bootstrapping T also leads to essentially the same C.I. though one does not need to know the polynomial $p(F, \cdot)$ for bootstrapping.

The idea of Theorem 2 can be stretched to obtain a two-step modified C.I. based on T_2. Let us assume that the coefficients of the polynomial $q(F, \cdot)$ are smooth functions of the initial few moments of F (in the sense of Remark 1). Assume also that F has sufficiently many moments, so that

$$T_2 = T_1 + n^{-1} q(F_n, T_1).$$

Define

$$b_1 = w_2 - p(F_n, w_2) n^{-1/2} + p(F_n, w_2) p'(F_n, w_2) n^{-1}$$

and

$$b_2 = w_1 - p(F_n, w_1) n^{-1/2} + p(F_n, w_1) p'(F_n, w_1) n^{-1}$$

where $w_1 = -z_{\alpha/2} - q(F_n, -z_{\alpha/2}) n^{-1}$ and $w_2 = z_{\alpha/2} - q(F_n, z_{\alpha/2}) n^{-1}$. Then arguments similar to those used for Theorem 2 can be used to show that

$$P(\theta_F < \hat{\theta}_n - b_1 v_n) = P(T_2 > z_{\alpha/2}) + o(n^{-1}) = \alpha/2 + o(n^{-1})$$

and

$$P(\theta_F > \hat{\theta}_n - b_2 v_n) = P(T_2 < -z_{\alpha/2}) + o(n^{-1}) = \alpha/2 + o(n^{-1}).$$

Thus $(\hat{\theta}_n - b_1 v_n, \hat{\theta} - b_2 v_n)$ is a two-step modified C.I. This interval leaves out $\alpha/2 + o(n^{-1})$ probability in each tail. Assuming that $p(F, \cdot)$ is an even polynomial, we present Table 1 containing information on the length and the coverage probabilities of the intervals based on T, T_1, and T_2. It is assumed that $n^{-1} r(F, \cdot)\phi(x)$ is the second term in the Edgeworth expansion of T.

Thus the first step modification basically distributes the error probability more evenly over the two sides than the naive one, keeping the same length and same coverage probability up to $O(n^{-1})$. The second stage modification provides even better balance for the error probabilities, though it may alter the length of the interval by a $O(1/n)$ amount. However, the increment (decrement) in the length is accompanied by an appropriate increment (decrement) in the total coverage probability.

TABLE 1

	C.I. (T)	C.I. (T_1)	C.I. (T_2)
Length	$2z_{\alpha/2}v_n$	$2z_{\alpha/2}v_n$	$2z_{\alpha/2}v_n + n^{-1}[q(F, -z_{\alpha/2}) - g(F, z_{\alpha/2})] + o(n-1)$
Lower side probability	$\alpha/2 + [p(F, -z_{\alpha/2})n^{-1/2} + n^{-1}r(F, -z_{\alpha/2})]\phi(x) + o(n^{-1})$	$\alpha/2 + n^{-1}q(F, -z_{\alpha/2})\phi(x) + o(n^{-1})$	$\alpha/2 + o(n^{-1})$
Upper side probability	$\alpha/2 - [p(F, z_{\alpha/2})n^{-1/2} + n^{-1}r(F, z_{\alpha/2})]\phi(x) + o(n^{-1})$	$\alpha/2 - n^{-1}q(F, z_{\alpha/2})\phi(x) + o(n^{-1})$	$\alpha/2 + o(n^{-1})$

How do we use the modified statistics for the purpose of testing a hypothesis on the parameter of interest θ_F? For the sake of argument, let us take $H_0{:}\theta_F = 0$ vs. $H_1{:}\theta_F > 0$. An obvious suggestion would be to compute T_1 or T_2 taking $\theta_F = 0$ and reject H_0 iff the computed value exceeds $100(1 - \alpha)$ percentile of the approximating distribution (α is the desired level of significance). It turns out that this straightforward way of using the modified statistics has a serious drawback. To understand this drawback, let us first note that T_1 typically has the following form:

$$T_1 = T + n^{-1/2}[\hat{C}_1(F) + \hat{C}_2(F)T^2]$$

where $C_1(F)$ and $C_2(F)$ depend upon certain moments of F (see the specific forms of T_1 in the following sections). Consider a case where T is of the form $\sqrt{n}(\hat{\theta}_F - \theta_F)/\hat{\sigma}(\theta_F)$ where $\operatorname{Var}(\hat{\theta}_F) \sim \sigma^2(\phi_F)/n$. Assume that $C_2(F) < 0$ (often, $\mu_3(T) \sim C_2(F)/\sqrt{n}$, where $\mu_3(T)$ stands for the third moment of T). If the true parameter θ_F is $> -\sigma(F)/C_2(F)$ and T is computed taking $\theta_F = 0$, it is clear that T_1 converges to $-\infty$, in probability. In other words, the power of the test described above converges to zero for all $\theta_F \in (-\sigma(F)/C_2(F), \infty)$! One way to avoid this highly undesirable phenomenon is to carry out the tests alternatively as follows: Construct a confidence interval for θ, one-sided or two-sided depending upon the problem, using the methods described earlier, and reject H_0 iff the confidence interval does not intersect the null hypothesis. Another possible way of avoiding the above phenomenon (at least in theory) is to reject H_0 right away if $T \geq c_n$, where c_n is a positive sequence $\to \infty$ s.t. $c_n = o(\sqrt{n})$ and $P_{H_0}(T \geq c_n) = o(n^{-1})$. If H_0 is not rejected, then compute T_1 and compare it with $100(1 - \alpha)$ percentile of the approximating distribution. At present, we do not see much practicality in the second idea, hence we recommend the use of confidence intervals for carrying out a test. Note that finer adjustments of the probabilities in each tail of a C.I. is specially important if the one-sided interval is to be used as the acceptance region of a one-sided test.

Hall (1983) considers $T_1(x) = T + P(F_n, x)/\sqrt{n}$ as a first step modified statistic *depending upon* x. Here typically $P(T + n^{-1/2}P(F_n, x) \leq x) = \Phi(x) + o(n^{-1/2})$. Now the Edgeworth expansion of $T_1(x)$ can be found for each x. This procedure can also be repeated to obtain $T_2(x)$, also a function of x. Since these modified statistics depend upon x, the remainder of the Edgeworth expansions are uniform only on compact sets, unlike the remainders of T_1 and T_2. In his Section 5, Withers (1983) uses the machinery he develops on Edgeworth expansions to obtain a statistic $\lambda(F_n, x)$, for a given sufficiently regular functional θ of a suitable F and a real number x, such that $P(\lambda(F_n, x) \leq \theta) = \Phi(x) + O(n^{-r})$, for any $r > 0$. Thus the result can be used to obtain a confidence interval with coverage probability $\alpha + O(n^{-r})$. Our approach of step-by-step modification of pivots and the use of bootstrap appears to be much more accessible.

In the following section, the case of the t-statistic for a general population is considered. The modifications t_1 and t_2 are explicitly discussed including the specific conditions that validate the various Edgeworth expansions involved. Also discussed is a problem relating to the bias of certain sample moments used in t_1 and t_2. The use of certain jackknifed versions of these statistics is suggested. Some simulation studies on the distribution of t_1 and t_2 are reported. Also included in this section is a Hodges-Lehmann deficiency result comparing t_2 and t_1^*, where t_1^* is the one-step corrected t-statistic on a symmetric population. One of the implications of this result is that "nothing is lost asymptotically" by using the two-step correction unnecessarily when the population is in fact normal. Section 3 deals with similar corrections in the case of random sampling without replacement. Specifically, the Studentized mean and the Studentized ratio estimators are examined. Section 4 deals with Studentized linear combinations of sample means from two or more independent samples. Finally, Section 5 contains the essentials of the proofs.

2. Modified t-statistic: t_1 and t_2. Let $X_1, X_2, \cdots, X_n$ be univariate observations from a population F whose mean is μ. Let σ^2, μ_3 and μ_4 denote the second, third and fourth central moments of F. Define $t = \sqrt{n}(\bar{X} - \mu)/\mathscr{S}_n$ where

$$\bar{X} = n^{-1} \textstyle\sum_1^n X_i \quad \text{and} \quad \mathscr{S}_n^2 = \sum_1^n (X_i - \bar{X})^2/(n-1).$$

The first-stage modification of t is given in the following theorem:

THEOREM 3. *Assume that F is continuous and has finite sixth moment. Define $t_1 = t + (\hat{\mu}_3/6\mathscr{S}_n^3\sqrt{n})[2t^2 + 1]$ where $\hat{\mu}_3 = n^{-1}\sum_1^n (X_i - \bar{X})^3$. Then,*

$$P(t_1 \leq x) = \Phi(x) + o(n^{-1/2})$$

uniformly in x.

We shall see later in this section that the remainder is $O(n^{-1})$ under stricter conditions. Theorem 3 is a consequence of Theorem 1 and the following Edgeworth expansion for t, which is valid under the conditions assumed:

$$P(t \leq x) = \Phi(x) + (\mu_3/6\sigma^3\sqrt{n})(2x^2 + 1)\phi(x) + o(n^{-1/2}).$$

The above Edgeworth expansion easily follows from Bhattacharya and Ghosh (1978). An outline of the proof is given in the last section.

In the following theorem, we define t_2 and state the conditions under which its distribution approximates the standard normal distribution up to a remainder of the order $o(n^{-1})$.

THEOREM 4. *Define*

$$t_2 = t_1 - \frac{b(F_n) - 6a(F_n)}{24n}(t_1^3 - 3t_1) - \frac{1}{2n}a(F_n)t_1$$

where

$$a(F) = \frac{1}{72\sigma^6}[-254\ \mu_3^2 + 168\ \mu_4\sigma^2 - 360\ \sigma^6]$$

and

$$b(F) = \frac{1}{6\sigma^6}[-191\ \mu_3^2 + 120\ \mu_4\sigma^2 - 252\ \sigma^6].$$

If F is differentiable in an open interval, where its derivative is strictly positive and $\int x^{12}\, dF$ *is finite, then*

$$P(t_2 \leq x) = \Phi(x) + o(n^{-1})$$

uniformly in x.

The differentiability condition on F can be replaced by a slightly weaker condition: F has an absolutely continuous component whose density is strictly positive on an open interval.

Here too, the main task is to show that t_1 admits an appropriate Edgeworth expansion (in this case the remainder is $o(n^{-1})$ and the $1/\sqrt{n}$ term is missing). One can immediately write down t_2 using this expansion and the ideas presented in the previous section.

That the bootstrap distribution of t_1 approximates its true distribution up to $o(n^{-1})$ is formally stated as

THEOREM 5. *Under the same conditions, as assumed in Theorem* 3,

$$n\ \sup_x |P^*(t_{1(B)} \leq x) - P(t_1 \leq x)| \to 0$$

where P^* *denotes probability under the empirical population* $\hat{F}$ *and* $t_{1(B)}$ *is the statistic* t_1 *computed on a random sample of size n from* $\hat{F}$ *and using* $\bar{X}$ *at the place of* μ.

Under the existence of 20 moments and the same differentiability condition on F, one can show that

$$n^{3/2}\sup_x |P^*(t_{2(B)} \leq x) - P(t_2 \leq x)| \to 0$$

where $P^*(t_{2(B)} \leq x)$ stands for the bootstrap distribution of t_2.

A question of asymptotic comparison arises in the use of the modified statistic t_2. One can specify the size of a test on μ up to $o(1/n)$ using t_2. But if the underlying population is *symmetric*, one can do the same thing with the one-step modification t_1^* of the t-statistic. Indeed, t_1^* differs from t by $O_p(n^{-1})$. Note that t_1^* uses the knowledge of symmetry whereas t_2 does not. The question then is:

How do t_2 and t_1^* compare on *symmetric populations* if both size and power are taken into account. This question is investigated in terms of the Hodges-Lehmann deficiency (a definition of the deficiency is contained in Section 5 where we have sketched a proof of the following theorem). It turns out that the deficiency is zero if the kurtosis μ_4/σ^4 of F equals 3, t_2, is deficient if $\mu_4/\sigma^4 > 3$ and t_1^* is deficient if $\mu_4/\sigma^4 < 3$. Thus t_2 and t_1^* are equivalent on a normal population. At present we do not have an heuristic explanation for this result, which could possibly suggest its generalization to other statistics.

The theorem also deals with the difference in the lengths $\lambda(t_2, \alpha)$ and $\lambda(t_1^*, \alpha)$ of confidence intervals based on t_2 and t_1^*. The coverage probabilities of both the intervals are $\alpha + o(1/n)$.

THEOREM 6 (a). *If F is symmetric and it satisfies the conditions of Theorem 4, then the Hodges-Lehmann Deficiency of t_2 as compared to t_1^* is given as*

$$\tfrac{1}{3}(\mu_4/\sigma^4 - 3)(1 + 2z_\alpha^2).$$

(b) *Let $\lambda(t_2, \alpha)$ and $\lambda(t_1^*, \alpha)$ be as mentioned above. Then,*

$$\lambda(t_2, \alpha) - \lambda(t_1^*, \alpha) = n^{-1}\left[\frac{2}{3} z_{\alpha/2}^3 + \frac{1}{3} z_{\alpha/2}\right]\left(\frac{\mu_4}{\sigma_4} - 3\right) S_n + o(n^{-1}).$$

A simulation study was carried out on the true distributions as well as on the bootstrap distributions of t, t_1 and t_2. The study was done on χ^2 populations with various degrees of freedom and sample sizes. It was found that the usual sample estimates of the coefficients of skewness and kurtosis used in computing t_1 and t_2 have substantial amount of bias. This tended to contaminate the improvement

TABLE 2
Simulation values for true and bootstrap distributions.

True values based upon 100,000 repetitions with sample size 30. The second entry is the median of 18 bootstrap values. Each bootstrap value is based upon 5,000 resamples. (All entries are in percentages.)

Distribution	Normal probabilities	t	t_1	t_2
χ_2^2	1	4.67, 3.27	1.67, 1.09	1.93, 1.52
	5	10.49, 8.45	5.76, 4.27	6.06, 5.09
	95	97.47, 96.61	94.77, 95.32	94.93, 94.52
	99	99.79, 99.53	98.95, 99.10	98.94, 98.64
χ_6^2	1	2.94, 2.36	1.53, 1.34	1.50, 1.54
	5	8.15, 6.98	5.62, 5.04	5.49, 5.50
	95	96.42, 95.90	94.62, 95.00	95.19, 94.60
	99	99.48, 99.34	98.82, 98.96	99.06, 98.70
χ_{10}^2	1	2.54, 2.08	1.53, 1.24	1.40, 1.56
	5	7.57, 6.68	5.71, 5.02	5.43, 5.60
	95	95.92, 95.44	94.49, 95.06	95.11, 94.72
	99	99.32, 99.18	98.72, 98.96	99.04, 98.66

achieved by t_1 and t_2. To avoid this problem we jackknifed the estimates before using them in computing t_1 and t_2. A two-stage jacknifed version of t_1 and t_2 we used in this simulation study; the jackknifed versions are denoted by $t_1(J_2)$ and $t_2(J_2)$. In general the improvement from t to t_1 was found to be substantial, while the further improvement from t_1 to t_2 was less substantial. For small sample sizes and large skewness, there were cases where t_1 and t_2 departed from normality by about the same amount or t_2 departed slightly more.

Table 2 contains t, $t_1(J_2)$ and $t_2(J_2)$ of sample size 30 for three χ^2 populations. We have tabulated the true distribution function and its bootstrap approximation at four normal percentiles. All the entries are in percentages. The first entry in each box is the true distribution based on 100,000 replications and the second one is the bootstrap approximation, median taken over 18 samples (5,000 bootstrap values for each sample). An appreciable amount of variability was found in the bootstrap approximations from sample to sample; however, the median value tended to beat the normal approximation, especially when the discrepancy between the true value and the normal approximation was substantial.

3. Modifications in the case of random sampling without replacement. Consider a finite population consisting of N distinguishable units $U_1, U_2, \cdots, U_N$. Let Y be a univariate characteristic whose value for U_i is Y_i. Let F_N denote the d.f. which assigns mass $1/N$ to each of Y_i's. The following notation is also used below:

$$\mu_N = N^{-1} \textstyle\sum_1^N Y_i, \quad \sigma_N^2 = N^{-1} \sum_1^N (Y_i - \mu_N)^2$$

and

$$\mu_{a,N} = N^{-1} \textstyle\sum_1^N (Y_i - \mu_N)^a \quad \text{for any real} \quad a > 2.$$

Suppose a random sample of size n is drawn from the population without replacement and the measurements on the sampled units are found to be $y_1, y_2, \cdots, y_n$. The following version of t-statistic is often used for the purpose of inference on μ_N:

$$t_s = \frac{\sqrt{n}(\bar{y} - \mu_N)}{(1-p)^{1/2}\mathcal{S}_n} \quad (t_s \equiv t\text{-statistics in sample surveys})$$

where $\bar{y} = (1/n) \sum_1^n y_i$, $\mathcal{S}_n^2\,[1/(n-1)] \sum_1^n (y_1 - \bar{y})^2$ and $p = (n/N)$. The classical CLT of Erdös and Renyi can be used to show that under certain conditions $P(t_s \le x) \to \Phi(x)$ as $(n, N - n) \to \infty$. In order to obtain the modified version of t_s, which eliminates the effect of population skewness, a one-term Edgeworth expansion for t_s is required. We borrow such a result from Babu and Singh (1982) (BS) to obtain $t_{s,1}$.

THEOREM 7. *If $n \to \infty$, $n/N \le 1/2$, $u_{6+\delta,N}$ is bounded for some $\delta > 0$ and F_n*

converges weakly to a continuous distribution, then

$$P(t_s \le s)$$
$$= \Phi(x) + \frac{1}{6\sqrt{n}} \frac{\mu_3, N}{\sigma_N^3} \left[3x^2 - \frac{1-2p}{1-p} (x^2 - 1) \right] (1-p)^{1/2} \phi(x) + o(n^{-1/2})$$

uniformly in x. Thus, if we define

$$t_{s,1} = t_s + \frac{1}{6\sqrt{n}} \frac{\hat{\mu}_{3,N}}{\hat{\sigma}_N^3} \left[3t_s^2 - \frac{1-2p}{1-p} (t_s^2 - 1) \right] (1-p)^{1/2},$$

then,

$$\sqrt{n} \sup_{x \in R} | P(t_{s,1} \le x) - \Phi(x) | \to 0.$$

For a bootstrap method which approximates the true distribution of t_s up to $o(n^{-1/2})$, see BS. It is not known to the authors if the condition F_N converges weakly to a smooth distribution is sufficient for the Edgeworth expansion of $t_{s,1}$ up to $o(n^{-1})$. The expansion would hold if we assume a superpopulation model, i.e., F_N is the empirical d.f. of a random sample of size N from a population F, where F satisfies the conditions of Theorem 4.

The formula for $t_{s,2}$ in terms of $t_{s,1}$ would be the same as that for t_2 in terms of t_1, provided we ignore the terms of order $O(1/N)$.

Finally, we consider in this section the ratio estimators used in sample surveys. Suppose, we have an auxiliary variable W, whose population values are W_1, W_2, $\cdots$, W_N and the sample values are w_1, w_2, $\cdots$, w_n. The ratio estimator of μ_N is defined as $\bar{W}r$ where

$$r = (\textstyle\sum_1^n y_i / \sum_1^n w_i) \quad \text{and} \quad \bar{W} = N^{-1} \sum_1^N W_i;$$

$\bar{W}$ is assumed known. The generally used pivotal version of the estimator is (recall that $p = n/N$)

$$r^* = \bar{W}\sqrt{n}(r - R)/[(1-p)n^{-1} \textstyle\sum_1^n (y_i - rw_i)^2]^{1/2}$$

where R is the population ratio $(\sum_1^N Y_i / \sum_1^N W_i)$. For the validity of one-term Edgeworth expansion of r^*, in view of the theory in BS, one needs to assume that the $6 + \delta$ absolute moment of the vector (Y, W) is bounded and also that the population satisfies *one* of the following two conditions:

(1) The joint population of (Y, W) has a weak limit which has a nonzero absolutely continuous component.

(2) The joint population of (Y, W) has a weak limit whose marginal corresponding to Y is continuous and whose marginal corresponding to Z is lattice.

Note that condition (2) above is suitable in the cases where the variable under study is continuous and the auxiliary information is some discrete measurement. For instance, in an agricultural experiment, Y could be the yield and W could be the number of times the plots are irrigated. Assuming the validity of the one-term Edgeworth expansion, the one-step correction of r^* is given as follows:

$$r_1^* = r^* - \frac{\sqrt{1-p}}{6v^3\sqrt{n}}\left\{[(r^*)^2 - 1]\frac{(1-2p)}{n(1-p)}\sum_1^n (y_i - rw_i)^3 - (r^*)^2\left[\frac{v^2}{n\bar{Z}}\sum_1^N (y_iw_i - rw_i^2) - \frac{1}{2n}\sum_1^n (y_i - rw_i)^3 - \frac{v^2}{n\bar{Z}}\sum_1^n (y_i - rw_i)(w_i - \bar{w})\right]\right\}$$

where $v^2 = (1/(n-1))\sum_1^n (y_i - rw_i)^2$.

Here too, the jackknifed versions $r_1^*(J_1)$ and $r_1^*(J_2)$ are expected to improve the performance of r_1^*. An appropriate way to bootstrap the one-step modified statistics in the situation of random sampling without replacement is to enlarge the sample $[N/n]$ times, repeating each sample value $[N/n]$ times, to form a sample population and resample from this population without replacement.

4. The k-sample mean problem. Suppose that k independent samples $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}$, $i = 1, 2, \cdots, k$ are drawn from the populations $F_1, F_2, \cdots, F_k$ having means $\mu_1, \mu_2, \cdots, \mu_k$. Let the linear combination $\sum_1^k \ell_i\mu_i$ be the parameter of interest. This situation corresponds to comparison of two population means or to inference on a contrast in the one-way ANOVA situation. In sample surveys, this situation occurs in the case of stratified sampling, in which case $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}$ is typically drawn without replacement from the ith stratum of the population.

Consider first Case 1 where $\{X_{i1}, \cdots, X_{in_i}\}$ are i.i.d. having distribution F_i. In this case, an appropriate version of t-statistic is as follows:

$$\bar{t} = \sum_1^k \ell_i(\bar{X}_i - \mu_i)/[\sum_1^k \ell_i^2 n_i^{-1} \mathscr{S}_i^2]^{1/2}$$

where $\bar{X}_i = n_i^{-1}\sum_1^{n_i} X_{ij}$ and $\mathscr{S}_i^2 = (n_i - 1)^{-1}\sum_1^{n_i}(X_{ij} - \bar{X}_i)^2$. Here the one-step modified statistic t_1 is

$$\bar{t}_1 = \bar{t} + (\hat{\bar{\mu}}_3/6\sqrt{n}(\hat{\bar{\sigma}})^3)(2\bar{t}^2 + 1)$$

where

$$n = \sum_1^k n_i, \quad \bar{\sigma}^2 = (1/n)\sum_1^k n_i\omega_i^2\sigma_i^2, \quad \omega_i = (\ell_i n/n_i),$$

$$\bar{\mu}_3 = \frac{1}{n}\sum_1^k n_i\omega_i^3\mu_{3,i}, \quad \sigma_i^2 = \int (x - \mu_i)^2\, dF_i \quad \text{and} \quad \mu_{3,i} = \int (x - \mu_i)^3\, dF_i.$$

The estimates $(\hat{\bar{\sigma}})^2$ and $\hat{\bar{\mu}}_3$ are taken to be $(1/n)\sum_1^k n_i\omega_i^2\mathscr{S}_i^2$ and $(1/n)\cdot\sum_1^k n_i\omega_i^3\hat{\mu}_{3,i}$ where $\hat{\mu}_{3,i} = (n_i)^{-1}\sum_1^{n_i}(X_{ij} - \bar{X}_i)^3$.

In case 2, consider the stratified sampling situation. Now the ith subsample is drawn from the finite population F_{i,N_i} without replacement. Let $p_i = n_i/N_i$ denote the sampling fraction. The stratified sampling version of above t and t_1 are given as follows: (let μ_{i,N_i} denote the mean of F_{i,N_i})

$$t_{st} = \sum_1^k \ell_i(\bar{X}_i - \mu_{i,N_i})/[\sum_1^k \ell_i^2(1 - p_i)\mathcal{S}_i^2/n_i]^2$$

$$t_{st,1} = t_{st} + (1/6\sqrt{n})(\hat{\bar{\sigma}})^3[3t_{st}^2\hat{\bar{\mu}}_3 - (t_{st}^2 - 1)\hat{\bar{\mu}}_3]$$

where

$$\bar{\sigma}^2 = \sum_1^k n_i(1 - p_i)\omega_i^2\sigma_i^2, \quad \bar{\mu}_3 = (1/n)\sum_1^k n_i(1 - p_i)^2\omega_i^3\mu_{3,i}$$

and

$$\hat{\bar{\mu}}_3 = (1/n)\sum_1^k n_i(1 - p_i)(1 - 2p_i)\omega_i^3\mu_{3,i}.$$

For technical reasons, we need to assume in the following theorem that $n/n_i \le \lambda < \infty$.

THEOREM 8. *In case 1, if $\int |x|^6 dF_i < \infty$ for each i, one of the F_i's is continuous, then*

$$\sqrt{n}\,\sup_x |P(\bar{t}_1 \le x) - \Phi(x)| \to 0.$$

In case 2, if $\int |x|^{6+\delta} dF_{i,N_i}$ is bounded for each i and one of the F_{i,N_i}'s converges weakly to a continuous d.f., then

$$\sqrt{n}\,|P(t_{st} \le x) - \Phi(x)| \to 0.$$

The essential arguments behind this theorem are outlined in Section 5. The second stage corrections in this set-up gets quite cumbersome and we chose to exclude that part. Of course, the asymptotic accuracy of 2nd stage correction would be achieved by bootstrapping t_1 or $t_{st,1}$, under suitable conditions.

We expect to return to a similar study of the general ANOVA set-up with unspecified error distribution.

5. The proofs.

PROOF OF THEOREM 1. Let us define $\bar{T}_1 = T + n^{-1}p(F, T)$. In view of the assumed Edgeworth expansion for T,

$$P(\bar{T}_1 \le x) = P(\{\bar{T}_1 \le x\} \cap \{|T| \le \log n\}) + o(n^{-1/2}).$$

On the set $|x| \le \log n$, $\eta(x) = x + n^{-1/2}p(F, x)$ is an increasing function of x, after certain n onwards. If $x \in [n(-\log n), \eta(\log n)]$, it follows that

$$P(\{\bar{T}_1 \le x\} \cap \{|T| \le \log n\})$$
$$= P(\{T \le x - n^{-1/2}p(F, x) + o(n^{-1/2})\} \cap \{|T| \le \log n\}) = \Phi(x) + o(n^{-1/2}).$$

Also, $\bar{T}_1 \notin [\eta(-\log n), \eta(\log n)]$ implies $\{|T| > \frac{1}{2}\log n\}$, after certain n, which has probability $o(n^{-1/2})$. Thus we have,

$$\sqrt{n}\,\sup_x |P(T + n^{-1/2}p(F, T) \le x) - \Phi(x)| \to 0.$$

Now the claim (3) follows using the condition (2).

REMARK 1. Under the conditions of the remark, it follows using an elementary argument that, on $\{|T| \le \log n\}$

$$\{|P(F, T) - p(F_n, T)| > \varepsilon\} \subseteq \cup_{i=1}^{r} \{|m_i - \mu_i'| > c/(\log n)^k\}$$

for some positive real c and k, where $m_i = n^{-1} \sum_1^n X_i^i$ and $\mu_i' = \int x^i \, dF$ (μ_r' is the highest order moment appearing in p). Now the conclusion follows from the following result.

If $\xi_1, \xi_2, \cdots, \xi_n$ are i.i.d. r.v.'s with mean zero and finite $(1.5 + \delta)$th moments for any $\delta > 0$, then

$$P(|\textstyle\sum_1^n \xi_i| > nc/(\log n)^k) = o(n^{-1/2}).$$

This result is obtained by trunking ξ_i's at $\sqrt{n}$ and applying Markov's inequality on the sum of truncated r.v.'s.

PROOF OF THEOREM 2. On a set with probability $1 - o(n^{-1/2})$, the function $\hat{\eta}(x) = x + n^{-1/2}p(F_n, x)$ is strictly increasing in x for $|x| \le \log n$, after certain n onwards. On this set,

$$\{\theta_F < \theta_n - a_1 v_n\} = \{T > a_1\} = \{T_1 > \hat{\eta}(a_1)\} = \{T_1 > z_{\alpha/2} + n^{-1}A_n\}$$

where $|A_n| \le c$ with probability $1 - o(n^{-1/2})$. The claim (4) follows this. One proves (5) similarly.

PROOF OF THEOREM 3. Using δ-method, we express t as follows:

$$t = \sqrt{n}\,[Z_1/\sigma - Z_1Z_2/2\sigma^3] + \gamma_n(t)$$

where $Z_1 = (\bar{X} - \mu)$ and $Z_2 = n^{-1} \sum_1^n (X_i - \mu)^2 - \sigma^2$ and, for every $\varepsilon > 0$, $P(|\gamma_n(t)| > \varepsilon n^{-1/2}) = o(n^{-1/2})$, so that $\gamma_n(t)$ does not matter for the one-term expansion. Now the Edgeworth expansion for $(\sqrt{n}/\sigma)[Z_1 - Z_1Z_2/2\sigma^2]$ follows from BG (Bhattacharya and Ghosh, 1978). As for the conditions, one requires the third absolute moment and the strong nonlatticeness of the vector $[(X - \mu), (X - \mu)^2]$. These conditions are easily concluded from the conditions of the theorem.

The formula for t_1 is deduced using Theorem 1.

PROOF OF THEOREM 4. Here too, the main task is to show that t_1 admits an appropriate Edgeworth expansion. One can immediately write down t_2 using this expansion and the ideas presented in the previous section. We express t_1 as

$$t_1 = \sqrt{n} g_n(Z_1, Z_2, Z_3) + \gamma_n^*(t_1)$$

where Z_1, Z_2 are as defined earlier and $Z_3 = [n^{-1} \sum_1^n (X_i - \mu)^3 - \mu_3]$,

$P(n\,|\,\gamma_n^*(t_1)\,| > \varepsilon) = o(n^{-1})$ and the function g_n is given as follows:

$$g_n(Z_1, Z_2, Z_3) = \left(\frac{1}{\sigma} - \frac{1}{\sigma n}\right)Z_1 - \frac{\mu_3}{4n\sigma^5} Z_2 + \frac{\sigma^4}{6n\sigma^7} Z_3$$

$$+ \frac{\mu_3}{6\sigma^3 n} + \frac{1}{6\sigma^5} [2\mu_3 Z_1^2 - 3\sigma^2 Z_1 Z_2]$$

$$+ \frac{1}{16\sigma^7} [-12\sigma^4 Z_1^3 - 20\mu_3 Z_1^2 Z_2 + 8\sigma^2 Z_1^2 Z_3 + 9\sigma^2 Z_1 Z_2^2].$$

Thus the problem reduces to the Edgeworth expansion of a trivariate mean. To this end, we appeal to Theorem 2(a) of BG. Except for the fact that the function g_n depends upon n, all other requirements of the above mentioned theorem is trivially seen to be satisfied. But the proof of this theorem of BG goes through for g_n without any change, using the fact that the coefficient of Z_1 in g_n is bounded away from zero. The following moment expansions are used in writing down the Edgeworth expansion explicitly:

$$E\sqrt{n}g_n = o(n^{-1}), \quad \mathrm{Var}(\sqrt{n}g_n) = 1 + a(F)/n + o(n^{-1}), \quad E(\sqrt{n}g_n)^3 = o(n^{-1}),$$

and

$$E(\sqrt{n}g_n)^4 = 3 + \frac{b(F)}{n} + o(n^{-1})$$

where $a(F)$ and $b(F)$ are as defined in the statement of Theorem 3.

PROOF OF THEOREM 5. An Edgeworth expansion up to $o(n^{-1})$ for $P^*(t_1^* \le x)$ is carried out, following the proof of the earlier theorem, step by step. The two statistics t_1^* and t_1 are structurally identical; the only difference lies in the underlying populations. In the case of t_1^* the population is $\hat{F}$, a discrete population!

It suffices to show that $\hat{F}$ induces all the required properties in F, asymptotically, with probability one. Clearly, in view of the SLLN, the first twelve moments of $\hat{F}$ are bounded, for all large n, a.s. The other property which $\hat{F}$ is required to possess is an appropriate limiting smoothness. In the expansion of t_1, the differentiability condition assumed on F is actually utilized to deduce that, for any $a > 0$,

$$\sup_{\varepsilon \le \|t\| \le n^a} |\,E \exp(it_1(X-\mu) + it_2(X-\mu)^2 + it_3(X-\mu)^3)\,| < 1$$

when $t = (t_1, t_2, t_3)$. We have the same bound for $Y - \bar{X}$, where Y is a random sample from $\hat{F}$, for all large n a.s. in view of the following:

LEMMA. *Let $\xi_1, \xi_2, \cdots, \xi_n$ be a random sample from a k-variate population, having a finite first absolute moment. Then, for any $a > 0$,*

$$\sup_{\|t\| \le n^a} |\,n^{-1} \textstyle\sum_1^n \exp(i \sum_1^k t \cdot \xi_i) - E \exp(it \cdot \xi_1)\,| \to 0 \text{ a.s.}$$

where $t = (t_1, t_2, \cdots, t_k)$.

PROOF. Let $R_n(t)$ denote $[n^{-1}\sum_1^n e^{it\cdot\xi_i} - E(e^{it\cdot\xi_1})]$. First note that, since $e^{it\cdot\xi_1}$ is bounded by 1 for all t, it follows using Markov's inequality that, for any $\varepsilon > 0$,

$$P(|R_n(t)| > \varepsilon) \le 4e^{-\delta n\varepsilon^2} \tag{5}$$

for a $\delta > 0$. Clearly, one can divide the zone $0 \le \|t\| \le n^a$ into small subzones, $\{I_j : j \in \Lambda\}$, each with diameter $\le \varepsilon$ s.t. the cardinality of Λ is $\le K(\varepsilon)n^{ka}$. Now, if $t_j \in I_j$ then

$$\sup_{\|t\|\le n^a}|R_n(t)| \le \max_{t_j : j\in\Lambda}|R_n(t_j)| + 2\varepsilon[E\|\xi_1\| + n^{-1}\textstyle\sum_1^n\|\xi_i\|]. \tag{6}$$

Using Bonferroni inequality and (5) it follows that

$$\max_{t_j : j\in\Lambda}|R_n(t_j)| \to 0 \quad \text{a.s.}$$

which implies the lemma, utilizing (6) and the fact that $E\|\xi_1\| < \infty$.

PROOF OF THEOREM 6(a). The definition of Hodges-Lehmann deficiency is briefly stated as follows: Consider a sequence of contiguous alternatives $\mu + \delta/\sqrt{n}$. Suppose, the tests t and t_1, each at a level α, require sample sizes n and $n + d_n$ respectively to attain a power β. (It is assumed that the sample sizes can be thought of as real variables, extending the power functions appropriately.) If $d_n \to d$ as $n \to \infty$, d is said to be the deficiency of t_1 relative to t.

To express the dependence on the sample size, let us write t_n for t and $t_{n,1}$ for t_1. Following the arguments of Theorem 4 and evaluating the first four cumulants of t_n under $\mathrm{H}_0{:}Ex = \mu$ and $\mathrm{H}_1{:}Ex = \mu + \delta/\sqrt{n}$ we deduce (let $K_4 = (\mu_4/\sigma^4 - 3)$)

$$P_{\mathrm{H}_0}(t_n \le x) = \Phi(x) + \frac{(K_4 - 3)}{12n}[x^3 - 3x]\phi(x) - \frac{x}{n}\phi(x) + o(n^{-1})$$

$$\begin{aligned} P_{\mathrm{H}_1}(t_n \le x) = \Phi(x - \delta) + \Bigg[&\frac{K_4 - 3}{12}((x-\delta)^3 - 3(x-\delta)) - \frac{3\delta}{4} \\ &- \frac{3\delta K_4}{8} - \frac{1}{2}\left(2 + \frac{\delta^2 K_4}{4} + \frac{\delta^2}{2}\right)(x - \delta) \\ &- \frac{\delta}{2}((x-\delta)^2 - 1)\Bigg]\frac{\phi(x-\delta)}{n} + o(n^{-1}). \end{aligned}$$

From these two expansions, we derive the expansion for the power of t_n, when the size is fixed at α:

$$\begin{aligned} &1 - \mathrm{Power}(t_1^*) \\ &= \Phi(z_\alpha - \delta) + \frac{\phi(z_\alpha - \delta)}{n}\left[\left(\frac{z_\alpha^2}{4} - \frac{K_4}{8} - \frac{K_4 z_\alpha^2}{4}\right)\delta + \frac{K_4 z_\alpha}{8}\delta^2 + \frac{K_4}{24}\delta^3\right] + o(n^{-1}). \end{aligned}$$

We repeat the procedure for $t_{n+d,1}$ under $\mathrm{H}_0{:}Ex = \mu$ and $\mathrm{H}_1{:}Ex = \mu + \delta/\sqrt{n}$. These expansions come out to be

$$P_{\mathrm{H}_0}(t_{n+d,1} \le x) = \Phi(x) - \left[\frac{1 + K_4}{4}(x^3 - 3x) + \frac{1}{2}\left(2 + \frac{7K_4}{3}\right)x\right]n^{-1}\phi(x) + o(n^{-1})$$

$$P_{H_1}(t_{n+d,1} \le x)$$
$$= \Phi\left(x - \delta - \frac{d}{2n}\right)$$
$$- \left[\frac{1+K_4}{4}((-\delta)^3 - 3(x-\delta)) + \frac{1}{2}\left(2 + \frac{\delta^2}{2} + \frac{7K_4}{3} + \frac{11\delta^2 K_4}{12}\right)(x-\delta)\right.$$
$$\left. + \frac{3\delta}{4} + \frac{25\delta K_4}{24} + \frac{\delta}{6}(3 + 4K_4)((x-\delta)^2 - 1)\right] n^{-1}\phi(x-\delta) + o(n^{-1}).$$

The power of $t_{n+d,1}$, when the size is α, is given as follows:

$$1\text{-Power}(t_2) = \Phi\left(z_\alpha - \delta - \frac{d}{2n}\right)$$
$$+ \left[\left(\frac{z_\alpha^2}{4} + \frac{K_4}{24} + \frac{K_4 z_\alpha^2}{12}\right)\delta + \frac{K_4 z_\alpha}{8}\delta^2 + \frac{K_4}{24}\delta^3\right] n^{-1}\phi(z_\alpha - \delta) + o(n^{-1}).$$

The expression $\frac{1}{3} K_4(1 + 2z_\alpha^2)$ of the deficiency follows from the above two power expansions.

PROOF OF THEOREM 6(b). It follows from the one-term Edgeworth expansions of t and t_1 for a symmetric population that

$$\lambda(t_2, \alpha) = 2z_{\alpha/2}v_n + \frac{2}{n}\left[\frac{1+K_4}{4}(z_{\alpha/2}^3 - 3z_{\alpha/2}) + \frac{1}{2}\left(2 + \frac{7K4}{3}\right)z_{\alpha/2}\right] + o(n^{-1})$$

and

$$\lambda(t_1^*, \alpha) = 2z_{\alpha/2}v_n + \frac{2}{n}\left[-\frac{K_4 - 3}{12}(z_{\alpha/2}^3 - 3z_{\alpha/2}) + z_{\alpha/2}\right] + o(n^{-1}).$$

The claim follows from these expansions.

The results of Section 3 including Theorem 7 are based on the Edgeworth expansions established in Babu and Singh (82).

PROOF OF THEOREM 8. The numerator of the k-sample t-statistic can be written as

$$\bar{Z}_1 = n^{-1} \textstyle\sum_{i=1}^k \sum_{j=1}^{n_i} \omega_i(X_{ij} - \mu_i)$$

where $\omega_i = n\ell_i/n_i$. Clearly

$$\bar{\sigma}^2 = \mathrm{Var}(\bar{Z}_1) = (n^{-1} \textstyle\sum_{i=1}^k n_i\omega_i^2\sigma_i^2)/n.$$

The 2-term stochastic expansion of $\bar{t}$ looks similar to that of t.

$$\bar{t} = \sqrt{n}\left[\frac{\bar{Z}_1}{\bar{\sigma}} - \frac{\bar{Z}_1\bar{Z}_2}{2\bar{\sigma}^3}\right] + o_p(n^{-1/2})$$

where $\bar{Z}_1$ and $\bar{\sigma}$ are as defined above and

$$\bar{Z}_2 = n^{-1} \textstyle\sum_{i=1}^k \sum_{j=1}^{n_i} \omega_i^2[(X_{ij}^2 - \mu_i)^2 - \sigma_i^2].$$

Thus the one-term Edgeworth expansion of $\bar{t}$ reduces to that of a bivariate mean consisting of k types of independent random variables. If it is assumed that for each i, $(n/n_i) \le \lambda < \infty$, i.e. the sample sizes n_i's are of the same order, then it suffices to assume the strong nonlatticeness of $[(X_{i1} - \mu_i), (X_{i1} - \mu_i)^2]$ for at least one i, which indeed is implied by the continuity of the corresponding population. The finiteness of the third absolute moment of the above vector is required for each i. In order to obtain explicit expression for the correction term, one needs the following moment expansions:

$$E(\bar{Z}_1 - \bar{Z}_1\bar{Z}_2/2\bar{\sigma}^2)(\sqrt{n}/\bar{\sigma}) = -(\bar{\mu}_3/2\sqrt{n}\bar{\sigma}^3) + o(n^{-1/2}),$$

$$E(\bar{Z}_1 - \bar{Z}_1\bar{Z}_2/2\bar{\sigma}^2)^2(n/\bar{\sigma}^2) = 1 + o(n^{-1/2})$$

and

$$E[(\bar{Z}_1 - (\bar{Z}_1\bar{Z}_2/2\bar{\sigma}^3))(\sqrt{n}/\bar{\sigma}) + (\bar{\mu}_3/2\sqrt{n}\bar{\sigma}^3)]^3 = -2(\bar{\mu}_3/\sqrt{n}\ \bar{\sigma}^3).$$

The arguments in the stratified sampling case are exactly similar.

REFERENCES

BABU, G. J. and SINGH, K. (1982). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.*, to appear.

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11**, 999–1003.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

HALL, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* **11** 569–576.

JOHNSON, N. J. (1978). Modified t-tests and confidence intervals for asymmetrical population. *J. Amer. Statist. Assoc.* **73** 536–544.

SINGH, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

WITHERS, C. S. (1983). Expansion for the distribution and quantiles of a regular functional of the empirical distribution with applications to non-parametric confidence intervals. *Ann. Statist.* **11** 577–587.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903