The Annals of Statistics
1982, Vol. 10, No. 2, 462–474

ON THE ASYMPTOTIC NORMALITY OF STATISTICS WITH ESTIMATED PARAMETERS

BY RONALD H. RANDLES

The University of Iowa and the University of Florida[1]

Often a statistic of interest would take the form of a member of a common family, except that some vital parameter is unknown and must be estimated. This paper describes methods for showing the asymptotic normality of such statistics with estimated parameters. Whether or not the limiting distribution is affected by the estimator is primarily a question of whether or not the limiting mean (derived by replacing the estimator by a mathematical variable) has a nonzero derivative with respect to that variable. Section 2 contains conditions yielding the asymptotic normality of U-statistics with estimated parameters. These results generalize previous theorems by Sukhatme (1958). As an example, we show the limiting normality of a resubstitution estimator of a correct classification probability when using Fisher's linear discriminant function. The results for U-statistics are extended to cover a broad class of families of statistics through the differential. Specifically, conditions are given which yield the asymptotic normality of adaptive L-statistics and an example due to de Wet and van Wyk (1979) is examined.

1. Introduction. Many complicated statistics used as estimators or test statistics would be members of some common family of statistics (for example, the U-statistic or L-statistic family) were it not for the fact that some vital parameters in the formulation of the statistics are unknown and hence are estimated with auxiliary statistics. Adaptive estimators are examples of ones which are often of this nature. Let $X_1, \cdots, X_n$ represent a random sample from some population and let

$$T_n(\hat{\lambda}) = T_n(X_1, \cdots, X_n; \hat{\lambda})$$

denote a statistic which is a function of the data $X_1, \cdots, X_n$ and also uses an estimator $\hat{\lambda}$ in an explicit way. The estimator $\hat{\lambda}$ also is a function of $X_1, \cdots, X_n$, consistently estimating the p-vector parameter λ. Suppose we replace the estimator $\hat{\lambda}$ in $T_n(\cdot)$ with a mathematical variable γ. Note that the true parameter value is still λ, but we are merely inserting the mathematical variable γ into the form of the statistic. In many settings the resulting statistic

$$T_n(\gamma) = T_n(X_1, \cdots, X_n; \gamma)$$

would be a member of a common family of statistics and hence would have appropriate theorems showing its asymptotic normality. It would be of interest to know whether or not the limiting distribution of $T_n(\hat{\lambda})$ is the same as that of $T_n(\lambda)$ and, if not, how the estimator $\hat{\lambda}$ affects the limiting distribution.

With the role of the estimator replaced by the mathematical variable γ, we find the limiting mean of the statistic $T_n(\gamma)$, namely

$$\mu(\gamma) = \lim_{n\to\infty} E_\lambda[T_n(\gamma)],$$

when the actual parameter value is λ. If the statistic $T_n(\gamma)$ is differentiably smooth as a

Received March 1980; revised November 1981.

[1] Current position.

AMS 1970 *subject classification.* Primary 62E20.

Key words and phrases. Asymptotic normality, estimated parameters, U-statistics, L-statistics, differentiable statistical functions.

function of γ at $\gamma = \lambda$, the Mean Value Theorem yields

$$\sqrt{n}[T_n(\hat{\lambda}) - \mu(\lambda)] = \sqrt{n}[T_n(\lambda) - \mu(\lambda)] + \sqrt{n}(\hat{\lambda} - \lambda)'\left[\frac{\partial}{\partial\gamma} T_n(\gamma)\Big|_{\gamma=\lambda^*}\right] \tag{1.1}$$

where the term on the far right involves the vector of partial derivatives of $T_n(\cdot)$ evaluated at a point λ^* between $\hat{\lambda}$ and λ. Slutsky's Theorem then can be used to show that

$$\begin{aligned}\sqrt{n}[T_n(\hat{\lambda}) - \mu(\lambda)] &= \sqrt{n}[T_n(\lambda) - \mu(\lambda)] \\ &\quad + \sqrt{n}(\hat{\lambda} - \lambda)' \lim_{n\to\infty} E\left[\frac{\partial}{\partial\gamma} T_n(\gamma)\Big|_{\gamma=\lambda}\right] + o_p(1).\end{aligned} \tag{1.2}$$

The asymptotic normality of (1.1) then follows from the joint limiting normality of

$$\sqrt{n}[T_n(\lambda) - \mu(\lambda), (\hat{\lambda} - \lambda)'].$$

To solidify the notation and also to illustrate the above expansion, consider the sample kth central moment

$$T_{\ln}(\bar{X}) = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^k$$

which is related to

$$T_{\ln}(\gamma) = \frac{1}{n}\sum_{i=1}^n (X_i - \gamma)^k.$$

For any fixed γ, this is an average of i.i.d. random variables to which the Central Limit Theorem would apply. It has mean function $\mu(\gamma) = E_\lambda[(X_1 - \gamma)^k]$, which takes on the value of the population kth central moment whenever γ equals the population mean λ. The sample kth central moment $T_n(\bar{X})$ is seen to be asymptotically normal using

$$\sqrt{n}[T_n(\bar{X}) - \mu(\lambda)] = \sqrt{n}[T_n(\lambda) - \mu(\lambda)] - k\sqrt{n}(\bar{X} - \lambda)\left[\frac{1}{n}\sum_{i=1}^n (X_i - \lambda^*)^{k-1}\right]$$

in the fashion indicated above, using the fact that

$$\frac{1}{n}\sum_{i=1}^n (X_i - \lambda^*)^{k-1} \to_p E[(X_1 - \lambda)^{k-1}]$$

where λ^* falls between $\bar{X}$ and λ.

In many situations, however, $T_n(\gamma)$ is not a differentiably smooth function of γ at $\gamma = \lambda$. In these cases, expansion (1.1) is inappropriate. Yet an expression like (1.2) may still be valid when the operations of differentiation and taking the limiting mean are reversed. That is, while the statistic is not sufficiently smooth as a function of γ, the limiting mean $\mu(\cdot)$ is. This motivates the use of

$$\sqrt{n}[T_n(\hat{\lambda}) - \mu(\lambda)] = \sqrt{n}[T_n(\lambda) - \mu(\lambda)] + \sqrt{n}(\hat{\lambda} - \lambda)'\left[\frac{\partial}{\partial\gamma}\mu(\gamma)\Big|_{\gamma=\lambda}\right] + o_p(1), \tag{1.3}$$

from which the limiting normality follows based on the joint limiting normality of $\sqrt{n}[T_n(\lambda) - \mu(\lambda), (\hat{\lambda} - \lambda)']$.

This paper presents conditions under which (1.3) holds. Because the assumptions required differ, we separate two cases. In Case A the limiting mean function $\mu(\cdot)$ has a zero differential at $\gamma = \lambda$. Here the asymptotic normality may be achieved by showing:

$$n^{1/2}[T_n(\hat{\lambda}) - T_n(\lambda)] \to_p 0 \tag{1.4}$$

and

$$n^{1/2}[T_n(\lambda) - \mu(\lambda)] \to_d N(0, \sigma^2(\lambda)). \tag{1.5}$$

Thus the limiting distribution of the standardized statistic with estimator $\hat{\lambda}$ is the same as though the actual parameter value λ were used. Examples of Case A settings where (1.4) and (1.5) were proved are given by Raghavachari (1965), Adichie (1974) and de Wet and van Wyk (1979).

If the limiting mean function $\mu(\gamma)$ has a nonzero differential at $\gamma = \lambda$, then the estimator $\hat{\lambda}$ may well affect the limiting normal distribution. These we call Case B situations. The asymptotic normality will then follow, provided we can show

$$n^{1/2}[T_n(\hat{\lambda}) - \mu(\hat{\lambda}) - T_n(\lambda) + \mu(\lambda)] \rightarrow_p 0 \tag{1.6}$$

and

$$n^{1/2}[T_n(\lambda) - \mu(\lambda) + \mu(\hat{\lambda}) - \mu(\lambda)] \rightarrow_d N(0, \tau^2). \tag{1.7}$$

Then by applying results like Theorem 3.3A of Serfling (1980) to (1.7) and, in addition, adding the left-hand sides of (1.6) and (1.7), we get that (1.3) holds and both sides of (1.3) have a limiting $N(0, \tau^2)$ distribution. Examples of Case B limiting distributions are given by Sukhatme (1958) and Gupta (1967), among others. Sukhatme obtains general results for steps (1.6) and (1.7) when $T_n(\cdot)$ is a U-statistic and the estimated parameter is a univariate location parameter. For settings in which one can obtain weak convergence of the conditional distribution of $T_n(\hat{\lambda})$ given $\hat{\lambda}$, Fligner and Hettmansperger (1979) provide an interesting alternative solution to the general problem. Loynes (1980) and Durbin (1975), among others, consider a related problem involving the distribution of statistics based on residuals. These formulations often use structural relationships between the observations and the parameters via, for instance, maximum likelihood estimates.

In this paper we provide conditions for showing the asymptotic normality of a common statistic with an estimated nuisance parameter. Section 2 contains theorems for U-statistics with estimated parameters which generalize results of Sukhatme (1958). The method is illustrated by showing the asymptotic normality of a resubstitution estimator for a correct classification probability when using Fisher's linear discrimination rule. Differentiable statistical functions are considered in Section 3. The differential, together with the results of Section 2, extend the method to a broad class of statistics with estimated parameters. This approach is then applied specifically to adaptive L-estimates in Section 4, yielding general conditions under which such statistics have limiting normal distributions. An example using an estimator of de Wet and van Wyk (1979) is also included.

Because it is more general, the thrust of this paper will emphasize Case B settings. However, the extra complication of Case B settings over corresponding Case A ones, generally requires slightly stronger assumptions. Conditions specifically designed for Case A or Case B settings will use an A or B in their index numbers.

2. *U*-statistics. We now consider random variables which would be U-statistics were it not for the fact that they contain an estimator. Let $X_1, \cdots, X_n$ denote a random sample from some population. Let $h(x_1, \cdots, x_r; \gamma)$ denote a symmetric kernel of degree r with expected value

$$\theta(\gamma) = E_\lambda[h(X_1, \cdots, X_r; \gamma)],$$

where λ denotes a parameter value. Both the kernel and its expected value may depend on γ. The corresponding U-statistic is then

$$U_n(\gamma) = \frac{1}{\binom{n}{r}} \sum_{\alpha \in A^*} h(X_{\alpha_1}, \cdots, X_{\alpha_r}; \gamma),$$

where A^* denotes the collection of all subsets of size r from the integers $\{1, 2, \cdots, n\}$. We would like to show the asymptotic normality of

$$n^{1/2}[U_n(\hat{\lambda}) - \theta(\lambda)], \tag{2.1}$$

but the standard U-statistic theorems do not apply because of the presence of the estimator

$\hat{\lambda}$ in the kernel. Hence we must apply the approach described in Section 1, the most difficult step of which is (1.6). The following conditons are useful in verifying this step.

CONDITION 2.2. Suppose

$$n^{1/2}(\hat{\lambda} - \lambda) = O_p(1).$$

CONDITION 2.3. Suppose there is a neighborhood of λ, call it $K(\lambda)$, and a constant $K_1 > 0$, such that if $\gamma \in K(\lambda)$ and $D(\gamma, d)$ is a sphere centered at γ with radius d satisfying $D(\gamma, d) \subset K(\lambda)$, then

(2.4) $$E[\sup_{\gamma' \in D(\gamma,d)} |h(X_1, \cdots, X_r; \gamma') - h(X_1, \cdots, X_r; \gamma)|] \leq K_1 d$$

and

(2.5) $$\lim_{d \to 0} E[\sup_{\gamma' \in D(\gamma,d)} |h(X_1, \cdots, X_r; \gamma') - h(X_1, \cdots, X_r; \gamma)|^2] = 0.$$

To verify this condition one must work directly with the supremum random variable. If the kernel value differences are bounded in a neighborhood of λ, the following lemma simplifies the effort. Its proof is direct.

LEMMA 2.6. *Suppose there exists an* $M_1 > 0$ *such that*

(2.7) $$|h(x_1, \cdots, x_r; \gamma) - h(x_1, \cdots, x_r; \lambda)| \leq M_1$$

for every $x_1, \cdots, x_r$ *and all* γ *in some neighborhood of* λ. *Then* (2.4) *implies* (2.5).

The important step (1.6), or (1.4), is then achieved by the following result, which generalizes a result due to Sukhatme (1958).

THEOREM 2.8. *Suppose Conditions* 2.2 *and* 2.3 *are satisfied. Then*

$$n^{1/2}[U_n(\hat{\lambda}) - \theta(\hat{\lambda}) - U_n(\lambda) + \theta(\lambda)] \to_p 0.$$

PROOF. Letting

$$Q_n(\mathbf{s}) = \frac{n^{1/2}}{\binom{n}{r}} \sum_{\alpha \in A^*} [\tilde{h}(X_{\alpha_1}, \cdots, X_{\alpha_r}; \lambda + n^{-1/2}\mathbf{s}) - \tilde{h}(X_{\alpha_1}, \cdots, X_{\alpha_r}; \lambda)],$$

where $\tilde{h}(\cdot; \gamma) = h(\cdot; \gamma) - \theta(\gamma)$, we need to show

$$Q_n(n^{1/2}(\hat{\lambda} - \lambda)) \to_p 0.$$

If $\varepsilon' > 0$ and $\delta' > 0$ are arbitrary, (2.2) shows that we can find a bounded sphere C in R^k centered at the origin such that

$$P[n^{1/2}(\hat{\lambda} - \lambda) \notin C] \leq \delta'/2$$

for every n. Note that

$$P[|Q_n(n^{1/2}(\hat{\lambda} - \lambda))| > \varepsilon'] \leq P[\sup_{\mathbf{s} \in C} |Q_n(\mathbf{s})| > \varepsilon'] + P[n^{1/2}(\hat{\lambda} - \lambda) \notin C].$$

Showing that

$$\sup_{\mathbf{s} \in C} |Q_n(\mathbf{s})| \to_p 0$$

follows the same steps as Theorem 3.1 of Sukhatme (1958). □

To achieve the asymptotic normality of (1.5) or (1.7), we specify the following conditions.

CONDITION 2.9A. Assume that $\theta(\gamma)$ has a zero differential at $\gamma = \lambda$, that

(2.10A) $$n^{1/2}[\hat{\lambda} - \lambda] = O_p(1)$$

and that

(2.11A) $$n^{1/2}[U_n(\lambda) - \theta(\lambda)] \to_d N(0, \tau^2),$$

where

(2.12A) $$\tau^2 = \mathrm{Var}(E[h(X_1, \cdots, X_r; \boldsymbol{\lambda}) \,|\, X_1]) > 0.$$

CONDITION 2.9B. Assume that $\theta(\gamma)$ has a nonzero differential at $\gamma = \boldsymbol{\lambda}$ and, in addition, that

(2.10B) $$n^{1/2}[U_n(\boldsymbol{\lambda}) - \theta(\boldsymbol{\lambda}), (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'] \to_d N_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Based on standard results, Condition 2.9B shows

(2.11B) $$n^{1/2}[U_n(\boldsymbol{\lambda}) - \theta(\boldsymbol{\lambda}) + \theta(\hat{\boldsymbol{\lambda}}) - \theta(\boldsymbol{\lambda})] \to_d N(0, \tau^2),$$

provided

(2.12B) $$\tau^2 = \mathbf{D}'\boldsymbol{\Sigma}\mathbf{D} > 0,$$

where

$$\mathbf{D}' = \left(1, \frac{\partial\theta(\cdot)}{\partial\gamma_1}, \cdots, \frac{\partial\theta(\cdot)}{\partial\gamma_p}\right)\bigg|_{\gamma=\boldsymbol{\lambda}}.$$

We can now state the desired asymptotic normality of a U-statistic with an estimator in the kernel.

THEOREM 2.13. *If Condition* 2.3 *holds and, in addition, one of Conditions* 2.9A *and* 2.9B *holds, then*

$$n^{1/2}[U_n(\hat{\boldsymbol{\lambda}}) - \theta(\boldsymbol{\lambda})] \to_d N(0, \tau^2)$$

provided $\tau^2 > 0$ *where* τ^2 *is given by* (2.12A) *or* (2.12B), *respectively*.

PROOF. The result follows immediately from Theorem 2.8 and Condition 2.9A or 2.9B. □

REMARKS 2.14. (i) Theorem 2.13 generalizes Theorem 4.1 of Sukhatme (1958) which dealt only with a bounded kernel and a λ which is a univariate location parameter.

(ii) Extension of Theorem 2.13 to a more general k-sample U-statistic is straightforward and will not be included here.

(iii) Note also that the results of these theorems remain valid if the observations $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are a random sample from some multivariate population.

(iv) Step (2.10B) is usually accomplished by approximating each component of $\hat{\boldsymbol{\lambda}}$ by an average of i.i.d. random variables, such that

$$n^{1/2}(\hat{\lambda}_j - \lambda_j) = n^{-1/2} \sum_{i=1}^n k_j(X_i) + o_p(1)$$

where $E[k_j(X_1)] = 0$. A differential often yields the appropriate $k_j(\cdot)$ (see Section 3). Step (2.11B) then follows from the fact that $\theta(\gamma)$ has a differential at $\gamma = \boldsymbol{\lambda}$, and the central limit theorem applied to

$$n^{-1/2} \sum_{i=1}^n \left[h_1(X_i; \boldsymbol{\lambda}) + \sum_{j=1}^p \left(\frac{\partial\theta(\cdot)}{\partial\gamma_j}\bigg|_{\gamma=\boldsymbol{\lambda}}\right) k_j(X_i)\right]$$

where

$$h_1(x; \gamma) = rE[\tilde{h}(X_1, \cdots, X_r; \gamma) \,|\, X_1 = x].$$

(v) If there exists an $M^* > 0$ such that

$$E[h^2(X_{i_1}, \cdots, X_{i_r}; \hat{\boldsymbol{\lambda}})] \le M^*$$

for every $1 \le i_1 \le \cdots \le i_r \le r$ and every $n = 1, 2, \cdots$, and if the conditions of Theorem 2.13 are satisfied, then the limiting distribution of the V-statistic

$$V_n(\hat{\boldsymbol{\lambda}}) = \frac{1}{n^r} \sum_{i_1=1}^n \cdots \sum_{i_r=1}^n h(X_{i_1}, \cdots, X_{i_r}; \hat{\boldsymbol{\lambda}})$$

is the same as that of the corresponding U-statistic $U_n(\hat{\lambda})$ in the conclusion of Theorem 2.13. This is easily proved by making minor modifications in Lemma 5.7.3 of Serfling (1980).

We conclude this section with an application to a discriminant analysis estimator. Suppose we wish to classify a univariate observation Z into one of two populations π_1 or π_2. The information available about these two populations comes from training samples: $X_1, \cdots, X_n$ i.i.d. $F(x)$, from π_1, and $Y_1, \cdots, Y_m$ i.i.d. $G(y)$, from π_2. We assume the distributions $F(\cdot)$ and $G(\cdot)$ have finite second moments and that $F(\cdot)$ is absolutely continuous with a density that is bounded above by $M^* > 0$ and is continuous in a neighborhood B^* of $(\mu_1 + \mu_2)/2$, the point halfway between the respective population means. Assuming equal prior probabilities and equal costs of misclassification, Fisher's linear discriminant function classifies $Z = z$ in π_1 if and only if

$$z \cdot \operatorname{sign}(\bar{x} - \bar{y}) \geq \tfrac{1}{2}(\bar{x} + \bar{y}).$$

The resubstitution estimator of the probabilty of correctly classifying an observation from π_1 is then $U_n(\bar{X}, \bar{Y})$, where

$$U_n(\gamma_1, \gamma_2) = \frac{1}{n} \sum_{i=1}^{n} \psi[X_i \cdot \operatorname{sign}(\gamma_1 - \gamma_2) - \tfrac{1}{2}(\gamma_1 + \gamma_2)]$$

and $\psi(t) = 1, 0$ as $t \geq, < 0$. We see that $U_n(\bar{X}, \bar{Y})$ is a one-sample U-statistic of degree one with estimators in its kernel. We assume that as $n \to \infty$, so does m in such a way that $n/m \to \delta^*$, for some $0 < \delta^*$. In addition, we note that $(\bar{X}, \bar{Y}) \to_p (\mu_1, \mu_2)$ which is assumed to satisfy $\mu_1 > \mu_2$. We now seek to apply Theorem 2.13 to this resubstitution estimator. First we note that for (γ_1, γ_2) in a small neighborhood of (μ_1, μ_2),

$$\theta(\gamma_1, \gamma_2) = 1 - F(\tfrac{1}{2}(\gamma_1 + \gamma_2)).$$

Thus both partial derivatives are continuous and equal to $-\frac{1}{2}f(\frac{1}{2}(\mu_1 + \mu_2))$ at $(\gamma_1, \gamma_2) = (\mu_1, \mu_2)$ and hence $\theta(\cdot)$ has a differential there. Let $K(\boldsymbol{\mu})$ be any neighborhood of (μ_1, μ_2) which does not include any points with equal coordinates and is such that $\gamma \in K(\boldsymbol{\mu})$ implies $\frac{1}{2}(\gamma_1 + \gamma_2) \in B^*$. Let $\gamma \in K(\boldsymbol{\mu})$ and $D(\gamma, d)$ denote a sphere of radius d centered at γ satisfying $D(\gamma, d) \subset K(\boldsymbol{\mu})$. Then

$$\sup_{\gamma' \in D(\gamma, d)} | h(\mathbf{x}; \gamma') - h(\mathbf{x}, \gamma) | = \begin{cases} 1 & \text{if } \frac{1}{2}(\gamma_1 + \gamma_2) - d < x < \frac{1}{2}(\gamma_1 + \gamma_2) + d, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the left-hand side of (2.4) is equal to

$$F(\tfrac{1}{2}(\gamma_1 + \gamma_2) + d) - F(\tfrac{1}{2}(\gamma_1 + \gamma_2) - d) \leq 2M^*d.$$

Using Lemma 2.6, we see that Condition 2.3 is satisfied. The joint normality of $\sqrt{n}\,[U_n(\mu_1, \mu_2) - \theta^*, \bar{X} - \mu_1, \bar{Y} - \mu_2]$ follows from the multivariate central limit theorem, where $\theta^* = \theta(\mu_1, \mu_2)$. Using Theorem 2.13, we see that

$$n^{1/2}[U_n(\bar{X}, \bar{Y}) - \theta^*] \to_d N(0, \tau^2)$$

where

$$\tau^2 = \theta^*(1 - \theta^*) - \theta^* f(\tfrac{1}{2}(\mu_1 + \mu_2))\{E[X_1 | X_1 > \tfrac{1}{2}(\mu_1 + \mu_2)] - \mu_1\} + 4^{-1}f^2(\tfrac{1}{2}(\mu_1 + \mu_2))(\sigma_1^2 + \delta^* \sigma_2^2)$$

and σ_i^2 is the variance of an observation from population π_i.

When the underlying populations are multivariate and the discriminant function is linear, Theorem 2.13 will yield asymptotic normality for a resubstitution estimator of a probability of correct classification, but the details of this general setting are more complicated.

3. Differentiable statistical functions. A convenient device for extending the results of Section 2 to a much broader class of statistics is provided by use of a differential.

The general approach described in this section is applied to adaptive L-estimates in Section 4. Using the differential as an extension device builds on the work of Boos (1977, 1979), Boos and Serfling (1980), and Serfling (1980), among others.

Let $T(F, \gamma)$ be a real-valued functional defined on $\mathscr{F} \times R^p$, where $\mathscr{F}$ denotes a convex class of distribution functions. Let $\mathscr{D}$ denote the linear space generated by differences $G - F$ of members of $\mathscr{F}$ and let $\|\cdot\|$ represent a norm on the space $\mathscr{D}$.

DEFINITION 3.1. The functional T is said to have a *differential* at (F, γ) with respect to $\|\cdot\|$, if there exists a function $T(F, \gamma; \Delta)$ defined on $\Delta \in \mathscr{D}$ and linear in the argument Δ, such that

$$T(G, \gamma) - T(F, \gamma) - T(F, \gamma; G - F) = o(\|G - F\|), \tag{3.2}$$

as $\|G - F\| \to 0$. Here $T(F, \gamma; \Delta)$ is called the differential. Moreover, T is said to be *uniformly differentiable* over γ in $B(\lambda)$, where $B(\lambda)$ is some neighborhood of λ in R^p, if

$$\sup_{\gamma \in B(\lambda)} | T(G, \gamma) - T(F, \gamma) - T(F, \gamma; G - F) | = o(\|G - F\|), \tag{3.3}$$

as $\|G - F\| \to 0$.

Let $F_n(\cdot)$ denote the empirical d.f. of a random sample $X_1, \cdots, X_n$ from a population with d.f. $F(\cdot)$. We are interested in the limiting distribution of $T(F_n, \hat{\lambda})$, where $\hat{\lambda}$ estimates λ in the sense that $n^{1/2}(\hat{\lambda} - \lambda) = O_p(1)$. Letting

$$\mu(\gamma) = T(F, \gamma),$$

we write

$$n^{1/2}[T(F_n, \hat{\lambda}) - \mu(\lambda)]$$

$$= n^{1/2}[T(F_n, \hat{\lambda}) - T(F, \hat{\lambda}) - T(F, \hat{\lambda}; F_n - F)] \tag{3.4a}$$

$$+ n^{1/2}[T(F, \hat{\lambda}; F_n - F) - T(F, \lambda; F_n - F)] \tag{3.4b}$$

$$+ n^{1/2}[T(F, \lambda; F_n - F) + \mu(\hat{\lambda}) - \mu(\lambda)]. \tag{3.4c}$$

To show the asymptotic normality of $n^{1/2}[T(F_n, \hat{\lambda}) - \mu(\lambda)]$, we will show that terms (3.4a) and (3.4b) both converge in probability to zero and that term (3.4c) has a limiting normal distribution.

Because the differential is linear in its third argument, we note that

$$T(F, \hat{\lambda}; F_n - F) = \frac{1}{n} \sum_{i=1}^{n} T(F, \hat{\lambda}; \delta_{X_i} - F), \tag{3.5}$$

where δ_x is the d.f. with point mass 1 at x. A similar expression holds with $\hat{\lambda}$ replaced by λ. Thus (3.5) is seen to be a U-statistic with an estimator $\hat{\lambda}$, corresponding to the kernel

$$h(x_1; \gamma) = T(F, \gamma; \delta_{x_1} - F),$$

which has been called the *influence curve*. In common cases when

$$\theta(\gamma) = E[h(X_1; \gamma)] = 0$$

and

$$\sigma^2(\gamma) = E[h^2(X_1; \gamma)] > 0,$$

the results of Section 2 apply.

In particular, we note that term (3.4b) is just the difference of two U-statistics, one with an estimator in it and the other of the same form but with the estimator replaced by the true parameter value to which it converges. Since the kernel has mean $\theta(\gamma) = 0$, Theorem 2.8 would apply to this difference, with the appropriate conditions placing certain restrictions on the form of the kernel and the underlying distribution. Examples are given in the next section.

The differential property is useful in showing that term (3.4a) converges in probability to zero. In fact, were it not for the presence of the estimator, term (3.4a) would be the type of term dealt with by Boos (1979) and Boos and Serfling (1980), for example. However, the estimator presents some additional complications which in some cases we can handle by showing uniform differentiability and by using the following lemma. Applications are found in Section 4.

LEMMA 3.6. *If $T(F, \gamma)$ is uniformly differentiable with respect to $\gamma \in B(\lambda)$ with differential $T(F, \gamma; \Delta)$, and if $n^{1/2}\| F_n - F\| = O_p(1)$ and $n^{1/2}(\hat{\lambda} - \lambda) = O_p(1)$, then*

$$n^{1/2}[T(F_n, \hat{\lambda}) - T(F, \hat{\lambda}) - T(F, \hat{\lambda}; F_n - F)] \to_p 0.$$

PROOF. Note that

$$P[n^{1/2}| T(F_n, \hat{\lambda}) - T(F, \hat{\lambda}) - T(F, \hat{\lambda}; F_n - F) | > \varepsilon]$$
$$\leq P\{n^{1/2} \sup_{\gamma \in B(\lambda)} | T(F_n, \gamma) - T(F, \gamma) - T(F, \gamma; F_n - F) | > \varepsilon$$
$$\text{and} \quad n^{1/2}\| F_n - F\| \leq M'\} + P\{n^{1/2}\| F_n - F\| > M'\} + P\{n^{1/2}(\hat{\lambda} - \lambda) \notin C\},$$

where C is a bounded sphere in R^p centered at the origin, $M' > 0$ and $B(\lambda) = \{\gamma \mid \gamma = \lambda + \mathbf{s}n^{-1/2}$ for some $\mathbf{s} \in C\}$. With proper choices for C and M', the last two terms are arbitrarily small and with n large enough the first term is zero. □

Finally, the asymptotic normality of term (3.4c) will follow under the conditions of the following lemmas, the proofs of which are immediate.

LEMMA 3.7A. *If $\mu(\gamma)$ has a zero differential at λ, and if $n^{1/2}(\hat{\lambda} - \lambda) = O_p(1)$, then*

$$n^{1/2}[T(F, \lambda; F_n - F) + \mu(\hat{\lambda}) - \mu(\lambda)] \to_d N(0, \sigma^2(\lambda)),$$

provided $\theta(\lambda) = E[T(F, \lambda; \delta_{X_1} - F)] = 0$ and $\sigma^2(\lambda) = E[T^2(F, \lambda; \delta_{X_1} - F)] > 0$.

LEMMA 3.7B. *If $\mu(\gamma)$ has a nonzero differential at λ and*

$$n^{1/2}[T(F, \lambda; F_n - F), (\hat{\lambda} - \lambda)'] \to_d N_{p+1}(\mathbf{0}, \mathbf{\Sigma})$$

then

$$n^{1/2}[T(F, \lambda; F_n - F) + \mu(\hat{\lambda}) - \mu(\lambda)] \to_d N(0, \tau^2),$$

provided

$$\tau^2 = \mathbf{D}'\mathbf{\Sigma}\mathbf{D} > 0,$$

where

$$\mathbf{D}' = \left(1, \frac{\partial\mu(\cdot)}{\partial\gamma_1}, \cdots, \frac{\partial\mu(\cdot)}{\partial\gamma_k}\right)\Bigg|_{\gamma=\lambda}.$$

REMARKS 3.8. (i). Expansion (3.4) is used rather than treating the differential of $T(F_n, \lambda(F_n))$ because it does not require $T(F_n, \gamma)$ to be differentiably smooth in γ.

(ii) Note that for the sup norm, the property $n^{1/2}\| F_n - F\|_\infty = O_p(1)$ follows from an inequality of Dvoretzky, Kiefer and Wolfowitz (1956).

4. Adaptive L-statistics. Again let $X_1, \cdots, X_n$ denote a random sample from a population with d.f. $F(\cdot)$, having an empirical d.f. $F_n(\cdot)$. A general functional representation of an L-statistic that depends on a parameter λ is

$$T(F, \lambda) = \textstyle\sum_{j=1}^{m} a_j(\lambda) T_j(F, \lambda), \tag{4.1}$$

where each term $T_j(F, \lambda)$ is one of the following three types.

TYPE I TERMS. $T_j(F, \lambda) = \int_0^1 F^{-1}(u) J_j(u, \lambda)\, du$, with

$$J_j(u, \lambda) = \begin{cases} 0 \text{ if } u < \alpha_j(\lambda) \\ 1 \text{ if } u \geq \alpha_j(\lambda). \end{cases}$$

TYPE II TERMS. $T_j(F, \lambda) = \int_0^1 F^{-1}(u) J_j(u, \lambda)\, du$, where $J_j(u, \lambda)$ is continuous in $u \in (0, 1)$.

TYPE III TERMS. $T_j(F, \lambda) = F^{-1}(p_j(\lambda))$.

In practice, the parameter λ is unknown and must be estimated from the data via $\hat{\lambda}$. For example, an adaptive symmetrically trimmed mean uses the functional form:

$$T(F_n, \hat{\lambda}) = \frac{1}{1 - 2\alpha(\lambda)} \int_{\alpha(\hat{\lambda})}^{1-\alpha(\hat{\lambda})} F_n^{-1}(u)\, du, \tag{4.2}$$

where the amount to be trimmed off each end, $\alpha(\hat{\lambda})$, is based on considerations such as tailweight of the population which, being unknown, must be estimated by the data. Thus the trimmed mean in (4.2) is described in form (4.1) as a sum of 2 Type I terms $\alpha(\gamma) = \alpha_1(\gamma) = 1 - \alpha_2(\gamma)$ and $a_1(\gamma) = -a_2(\gamma) = (1 - 2\alpha(\gamma))^{-1}$. Adaptive trimmed means of this form have been proposed by Jaeckel (1971) and by de Wet and van Wyk (1979). The particular class investigated by de Wet and van Wyk includes one which uses

$$\alpha(\gamma) = \begin{cases} .05 & \text{if } \gamma \leq 1.75, \\ 2(\gamma - 1.75) + .05 & \text{if } 1.75 < \gamma < 1.95, \\ .45 & \text{if } \gamma \geq 1.95, \end{cases} \tag{4.3}$$

where $\hat{\lambda}$ measures tailweight via

$$\hat{\lambda} = \frac{\bar{U}(.20) - \bar{L}(.20)}{\bar{U}(.50) - \bar{L}(.50)} \tag{4.4}$$

with $\bar{U}(\beta)$ and $\bar{L}(\beta)$ respectively denoting the averages of the upper and lower $n\beta$ order statistics (fractional items are included, if $n\beta$ is not an integer); see, for example, Hogg (1974) for a discussion of this tailweight measure. So when the data indicates a heavy- (light-) tailed distribution, an α-trimmed mean is used with a large (small) α.

Thus we are investigating L-statistics which have a form depending on characteristics λ of the underlying distribution. When these characteristics are estimated from the data via $\hat{\lambda}$, the L-statistic adapts to the underlying distribution in the sense described by Hogg (1967, 1974). Here we are concerned with showing the asymptotic normality of

$$n^{1/2}[T(F_n, \hat{\lambda}) - T(F, \lambda)].$$

The method used is that of Section 3; in particular, we consider the steps associated with expansion (3.4). Because of the additive nature of (4.1), it suffices to handle separately the individual $T_j(\cdot)$'s in the sum (4.1) when showing that (3.4a) and (3.4b) converge to zero in probability; see Remark 4.16(ii). When showing the asymptotic normality of the term (3.4c), we must consider the L-statistic functional as a whole, as will be illustrated with an example. First we specify conditions for each of the three types of terms $T_j(\cdot)$ under which (3.4a) and (3.4b) converge to zero in probability. We begin with Type I terms.

CONDITION 4.5. Suppose $\alpha(\cdot)$ satisfies $0 < \alpha(\lambda) < 1$ and

$$|\alpha(\gamma') - \alpha(\gamma)| \leq M|\gamma - \gamma'|$$

for some $M > 0$ and every γ and γ' in a neighborhood of λ. Suppose also that the distribution $F(\cdot)$ is absolutely continuous in some neighborhood of $F^{-1}(\alpha(\lambda))$ with a density $f(\cdot)$ that is bounded away from zero in that neighborhood. In addition, we assume that $F(\cdot)$ satisfies $\int |x|\, dF(x) < \infty$ and $n^{1/2}(\hat{\lambda} - \lambda) = O_p(1)$.

The first step is to show that Lemma 3.6 is satisfied for the sup norm, and hence the corresponding (3.4a) term converges to zero in probability. To do that, we demonstrate uniform differentiability (3.3). Using the notation and approach of Boos (1979), we note that

$$
\begin{aligned}
(4.6)\qquad |T(G,\gamma) - T(F,\gamma) - T(F,\gamma; G-F)| &\le \|G - F\|_\infty \int |W_{G,F,\gamma}(x)|\,dx \\
&= \|G-F\|_\infty \int_{Q(x,\gamma)} \frac{|G(x)-\alpha(\gamma)|}{|G(x)-F(x)|}\,dx \\
&\le \|G-F\|_\infty \int_{Q(x,\gamma)} dx,
\end{aligned}
$$

where

$$Q(x,\gamma) = \{x \mid G(x) \ge \alpha(\gamma) > F(x) \quad \text{or} \quad F(x) \ge \alpha(\gamma) > G(x)\},$$

$$\|H\|_\infty = \sup_x |H(x)|,$$

and

$$(4.7)\qquad T(F,\gamma;\Delta) = -\int_{-\infty}^{\infty} \Delta(t) J(F(t),\gamma)\,dt.$$

Using Condition 4.5, we select a neighborhood of λ, denoted $B(\lambda)$, such that $0 < \alpha_L \le \alpha(\gamma) \le \alpha_U < 1$ for every $\gamma \in B(\lambda)$ and $f(x) \ge M^* > 0$ for every x satisfying $F^{-1}(\alpha_L) \le x \le F^{-1}(\alpha_U)$. Then for $\|G - F\|_\infty$ sufficiently small,

$$\int_{Q(x,\gamma)} dx \le \frac{2\|G-F\|_\infty}{M^*}$$

for every $\gamma \in B(\lambda)$. Thus we have demonstrated uniform differentiability. Using Condition 4.5 and Remark 3.8ii, Lemma 3.6 shows that the corresponding term (3.4a) converges to zero in probability.

The second step of the process of handling the Type I terms is to show the corresponding term (3.4b) converges to zero in probability by applying Theorem 2.8. This is easily accomplished by writing out the nature of the kernel and considering Condition 2.3, using Lemma 2.6 to verify (2.5).

Next we consider Type II terms which have a smooth $J(u,\gamma)$ function. In the regularity conditions which follow, $q(u)$ represents a nonnegative, real-valued function satisfying $0 < q(u) < M^*$ for $0 < u < 1$, which is used to construct a q-norm, $\|h(x)\|_{q(F)} = \|h(x)/q(F(x))\|_\infty$.

CONDITION 4.8. Suppose there is some fixed neighborhood $B(\lambda)$ of λ, such that $J(u,\gamma)$ is uniformly continuous uniformly in $B(\lambda)$, in the sense that, for every $\varepsilon > 0$, there is a $\delta > 0$ such that $0 \le u_1 \le u_2 \le 1$ and $|u_1 - u_2| < \delta$ implies

$$(4.9)\qquad |J(u_1,\gamma) - J(u_2,\gamma)| < \varepsilon$$

for every $\gamma \in B(\lambda)$. Also, assume there is some $M > 0$ for which $|J(u,\gamma)| \le M$, for every $0 \le u \le 1$ and $\gamma \in B(\lambda)$. In addition, suppose that $n^{1/2}(\hat\lambda - \lambda) = O_p(1)$, that

$$(4.10)\qquad n^{1/2}\|F_n - F\|_{q(F)} = O_p(1),$$

$$(4.11)\qquad \int q(F(x))\,dx < \infty,$$

and that

$$(4.12)\qquad v(\gamma',\gamma) = \left\| \frac{J(\cdot,\gamma') - J(\cdot,\gamma)}{q(\cdot)} \right\|_\infty$$

satisfies

$$v(\gamma', \gamma) \le M|\gamma - \gamma'|$$

for some $M > 0$ and every γ and γ' in some neighborhood of λ.

Under Condition 4.8, we verify that corresponding terms (3.4a) and (3.4b) converge to zero in probability in much the same fashion as with Type I terms. With Type II terms the role of the differential is again played by (4.7). We write out term (3.4a) as was done in (4.6) and work with an upper bound to verify the conditions of Lemma 3.6. We then apply Lemma 2.6 and Theorem 2.8 to show the term (3.4b) converges to zero in probability.

For Type III terms, the role of the differential is played by

$$T(F, \gamma; F_n - F) = [p(\gamma) - F_n(F^{-1}(p(\gamma)))][f(F^{-1}(p(\gamma)))]^{-1}.$$

We will assume the following.

CONDITION 4.13. Suppose $p(\gamma)$ has a differential at λ. In addition, suppose $F(x)$ is absolutely continuous and differentiable in some neighborhood of $F^{-1}(p(\lambda))$. Assume a density $f(x)$ is bounded away from zero and is continuous in that same neighborhood. Also, suppose $n^{1/2}(\hat{\lambda} - \lambda) = 0_p(1)$.

Rather than use the differential approach, we will combine the terms in (3.4a) and (3.4b) and show that this sum converges in probability to zero.

LEMMA 4.14. *Under Condition* 4.13,

$$n^{1/2}\{F_n^{-1}(\hat{p}) - F^{-1}(\hat{p}) - [p - F_n(F^{-1}(p))][f(F^{-1}(p))]^{-1}\} \to_p 0,$$

where $p = p(\lambda)$ *and* $\hat{p} = p(\hat{\lambda})$.

PROOF. The result will follow provided we can show

$$n^{1/2}\{F_n^{-1}(\hat{p}) - F^{-1}(p) - [\hat{p} - p][f(F^{-1}(p))]^{-1} - [p - F_n(F^{-1}(p))][f(F^{-1}(p))]^{-1}\} \to_p 0.$$

Define

$$V_n(s) = n^{1/2}\{F_n^{-1}(p + n^{-1/2}s) - F^{-1}(p) - n^{-1/2}s[f(F^{-1}(p))]^{-1}\}$$

$$Z_{t,n}(s) = n^{1/2}\{F(F^{-1}(p) + n^{-1/2}s[f(F^{-1}(p))]^{-1} + n^{-1/2}t) - F_n(F^{-1}(p) + n^{-1/2}s[f(F^{-1}(p))]^{-1} + n^{-1/2}t)\}\{f(F^{-1}(p))\}^{-1}$$

and

$$W_n = n^{1/2}\{[p - F_n(F^{-1}(p))][f(F^{-1}(p))]^{-1}\}.$$

For arbitrary $\varepsilon > 0$ and $\delta > 0$, $\sqrt{n}\ (\hat{p} - p) = 0_p(1)$ implies there is a $c > 0$ such that for every n,

$$P\{\sqrt{n}|\hat{p} - p| > c\} < \delta/2.$$

Since

$$P\{|V_n(\sqrt{n}(\hat{p} - p)) - W_n| > \varepsilon\} \le P\{\sqrt{n}|\hat{p} - p| > c\} + P\{\sup_{|s|\le c}|V_n(s) - W_n| > \varepsilon\},$$

we need to show

$$\sup_{|s|\le c}|V_n(s) - W_n| \to_p 0. \tag{4.15}$$

This proof follows the same basic steps as the proof of Theorem 1 in Ghosh (1971). The major difference being that in the process we must show

$$\sup_{|s|\le c}|Z_{t,n}(s) - W_n| \to_p 0,$$

which follows in the same fashion as the proof of Theorem 2.8. □

To show that the term corresponding to (3.4b) converges to zero in probability, the procedure is similar to the other two cases, in that Theorem 2.8 is applied via Lemma 2.6.

REMARK 4.16. (i) The asymptotic normal distribution associated with the term (3.4c) for an L-statistic $T(F, \lambda)$ given by (4.1) is found by dealing with $T(F, \lambda)$ as a whole. That is, we show the asymptotic normality of

$$n^{1/2}[T(F, \lambda; F_n - F) + \mu(\hat{\lambda}) - \mu(\lambda)] \tag{4.17}$$

via Lemma 3.7A or 3.7B.

(ii) In showing that (3.4a) and (3.4b) converge to zero in probability for the individual terms $T_j(F, \gamma)$ of the L-statistic described in (4.1), we ignored the multiplicative terms $a_j(\gamma)$. Examination of the steps (3.4a) and (3.4b) shows that these quantities will not affect the convergence provided

$$n^{1/2}[a_j(\hat{\lambda}) - a_j(\lambda)] = O_p(1), \qquad j = 1, \cdots, m.$$

However, these multiplicative terms will affect step (4.17), for which we must show

$$n^{1/2}\{\textstyle\sum_{j=1}^m a_j(\lambda) T_j(F, \lambda; F_n - F) + \sum_{j=1}^m [a_j(\hat{\lambda})\mu_j(\hat{\lambda}) - a_j(\lambda)\mu_j(\lambda)]\}$$

has a limiting normal distribution.

We conclude this section by applying its results to the adaptive trimmed mean described in (4.2) through (4.4). In their paper, de Wet and van Wyk confined attention to distributions symmetric about some value μ_*. Since they also trimmed symmetrically, their functional then satisfied

$$\mu(\gamma) = T(F, \gamma) = \mu_*$$

for every γ. Thus $\mu'(\gamma) = 0$ and, as de Wet and van Wyk demonstrated, their adaptive estimator had the same limiting normal distribution as though they had used a fixed amount of trimming $\alpha(\lambda)$ where λ is the value which satisfies $n^{1/2}(\hat{\lambda} - \lambda) = O_p(1)$ for $\hat{\lambda}$ defined in (4.4). To illustrate the application of results in this section, we will assume $1.75 < \lambda < 1.95$ and find the limiting distribution of the de Wet-van Wyk estimator when the underlying population is not necessarily symmetric. We see from Condition 4.5 that we must assume $\int |x| \, dF(x) < \infty$ and that $F(\cdot)$ is absolutely continuous with a density bounded away from zero in some neighborhood of $F^{-1}(\alpha(\lambda))$ and $F^{-1}(1 - \alpha(\lambda))$. From (3.4c) we note that the limiting distribution of

$$n^{1/2}[T(F_n, \alpha(\hat{\lambda})) - \mu(\alpha(\lambda))]$$

is the same as that of

$$n^{1/2}[T(F_n, \alpha(\lambda)) - \mu(\alpha(\lambda)) + \mu(\alpha(\hat{\lambda})) - \mu(\alpha(\lambda))].$$

We also note that both $\mu(\cdot)$ and $\alpha(\cdot)$ are uniformly differentiable in a neighborhood of $\alpha = \alpha(\lambda)$ with

$$\mu'(\alpha) = \frac{1}{(1 - 2\alpha)} \{2\mu(\alpha) - [F^{-1}(\alpha) + F^{-1}(1 - \alpha)]\}.$$

In addition, we see that $\hat{\lambda}$ is the ratio of two L-statistics, i.e.,

$$\hat{\lambda} = \frac{\int F_n^{-1}(u) J_N(u) \, du}{\int F_n^{-1}(u) J_D(u) \, du}$$

where

$$J_N(u) = \begin{cases} 5 \text{ if } 0 < u < .2, \\ -5 \text{ if } .8 < u < 1, \\ 0 \text{ otherwise,} \end{cases}$$

and

$$J_D(u) = \begin{cases} 2 \text{ if } 0 < u < .5, \\ -2 \text{ if } .5 < u < 1, \\ 0 \text{ otherwise.} \end{cases}$$

Letting

$$\mu_N = \int F^{-1}(u) J_N(u)\, du \quad \text{and} \quad \mu_D = \int F^{-1}(u) J_N(u)\, du$$

and applying the form of an L-statistic differential found in (4.7), we see that the limiting distribution is the same as that of an L-statistic with $J(\cdot)$ function

$$J_*(u) = J(u, \alpha) + \frac{2\mu'(\alpha)}{\mu_D} J_N(u) - \frac{2\mu'(\alpha)\mu_N}{\mu_D^2} J_D(u)$$

where $\alpha \equiv \alpha(\lambda)$. This is known to be normal—see, for example, the corollary in Section 4 of Boos (1979)—provided its variance ((3.7) of Boos) is positive and if, in addition to the previous assumptions, $F(\cdot)$ is continuous at $F^{-1}(.2)$, $F^{-1}(.5)$ and $F^{-1}(.8)$ and satisfies

$$\int q(F(x))\, dx < \infty$$

for some $q(\cdot)$ in Q_1 of Boos.

Acknowledgements. The author would like to thank T. de Wet, R. V. Hogg, H. Iverson, the editor, associate editor and referee for helpful suggestions.

REFERENCES

ADICHIE, J. N. (1974). Rank score comparison of several regression parameters. *Ann. Statist.* **2** 396–402.

BERK, R. H. (1970). Consistency a posteriori. *Ann. Math. Statist.* **41** 894–907.

BOOS, D. D. (1977). The differential approach in statistical theory and robust inference. Unpublished dissertation, Florida State Univ.

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959.

BOOS, D. D. and SERFLING, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions, with applications to M-estimates. *Ann. Statist.* **8** 618–624.

DE WET, T. and VAN WYK, J. W. J. (1979). Some large sample properties of Hogg's adaptive trimmed means. *South African Statist. J.* **13** 53–69.

DURBIN, J. (1973). Tests of model specification based on residuals. In *A Survey of Statistical Design and Linear Models* (ed. J. N. Srivastava), North Holland, Amsterdam.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

FLIGNER, M. A. and HETTMANSPERGER, T. P. (1979). On the use of conditional asymptotic normality. *J. Roy. Statist. Soc. B* **41** 178–183.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

GUPTA, M. K. (1967). Asymptotically nonparametric tests of symmetry. *Ann. Math. Statist.* **38** 849–866.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOGG, R. V. (1967). Some observations on robust estimation. *J. Amer. Statist. Assoc.* **62** 1179–1186.

HOGG, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.* **69** 909–923.

JAECKEL, L. A. (1971). Some flexible estimates of location. *Ann. Math. Statist.* **42** 1540–1552.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

RAGHAVACHARI, M. (1965). The two-sample scale problem when locations are unknown. *Ann. Math. Statist.* **36** 1236–1242.

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

SUKHATME, B. V. (1958). Testing the hypotheses that two populations differ only in location. *Ann. Math. Statist.* **29** 60–78.

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FL 32611