LIMITING SECOND ORDER DISTRIBUTIONS FOR FIRST ORDER FUNCTIONALS, WITH APPLICATION to L- AND M-STATISTICS

by

Dennis D. Boos

Department of Statistics
North Carolina State University

Institute of Statistics Mimeo Series # 1152

December, 1977

SUMMARY

LIMITING SECOND ORDER DISTRIBUTIONS FOR FIRST ORDER FUNCTIONALS, WITH APPLICATION TO L- AND M-STATISTICS

Let $T(\cdot)$ be a real-valued functional defined on the space of distribution functions (d.f.'s) and $T(F_n)$ the statistic obtained by evaluating $T(\cdot)$ at F_n, the usual empirical d.f. for a sample $X_1, \ldots, X_n$. Under restrictions on the X_i, existence of a nonzero (first) differential for T leads to the approximation $\sqrt{n}\ (T(F_n) - T(F) - n^{-1}\ \Sigma V(X_i)) \xrightarrow{p} 0$ and to asymptotic normality of $T(F_n)$. Existence of a second differential for T leads to the approximation $n(T(F_n) - T(F) - n^{-1}\Sigma V(X_i) - n^{-2}\Sigma\Sigma Q(X_i X_j)) \xrightarrow{p} 0$, which then yields the limiting distribution of $n(T(F_n) - T(F) - n^{-1}\Sigma V(X_i))$. The variance of this latter asymptotic distribution is used to compare estimators with identical first order approximating sums $n^{-1}\Sigma V(X_i)$. Application is provided by L- and M-statistics with special attention given to α-trimmed means and "Hubers."

1. Introduction. The study of statistical functions (real-valued functionals defined on the space of distribution functions (d.f.'s)) was pioneered by von Mises (1947) and continued by Kallianpur and Rao (1955), Fillippova (1962), Kallianpur (1963), Gregory (1976), and Boos and Serfling (1977). Most of these efforts have concentrated on finding the limiting distribution of $n^{k/2}(T(F_n) - T(F))$, where T is the functional of interest, F_n is the usual empirical d.f. generated by a sample $X_1, \ldots, X_n$ from a distribution F, and k refers to the "order" of the functional.

If $T(\cdot)$ has a nonvanishing generalized first derivative $T(F;\Delta)$, then k=1 and under suitable regularity conditions on the sequence of r.v.'s $\{X_i\}$ from F, it follows that

$$(1.1) \qquad \sqrt{n}\ (T(F_n) - T(F) - \frac{1}{n}\sum_{i=1}^{n} V(X_i)) \xrightarrow{p} 0\ ,\ n\to\infty\ ,$$

where $V(x) = T(F;\ \delta_x - F)$ and δ_x refers to the d.f. degenerate at x. In other contexts $T(F;\delta_x - F)$ is called the influence curve of T (see Hampel (1974)). We will use $\xrightarrow{p}$ and $\xrightarrow{d}$ to refer respectively to convergence in probability and convergence in distribution. Clearly, (1.1) provides an approximation to the error $T(F_n) - T(F)$ which yields asymptotic normality of $T(F_n)$ under the restrictions of appropriate central limit theory. Many common statistical functions such as the mean, $T(F) = \int x dF$, the p^{th} quantile, $T(F) = F^{-1}(p)$, and the variance, $T(F) = \int [x - \int x dF]^2 dF$, satisfy (1.1). A rich source of location functionals is Andrews, et al. (1972). In addition to asymptotic normality, Boos and Serfling (1977) provide a law of the iterated logarithm for first order functionals.

If the first nonvanishing derivative is the second, designated $T^2(F;\Delta_1,\Delta_2)$, then under suitable regularity conditions

$$(1.2) \qquad n(T(F_n) - T(F) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q(X_i,X_j)) \xrightarrow{p} 0\ ,\ n\to\infty\ ,$$

where $2Q(x,y) = T^2(F;\delta_x - F, \delta_y - F)$. Under further restrictions we then have $n(T(F_n) - T(F)) \xrightarrow{d} \Sigma_{i=1}^{\infty}\Sigma_{j=1}^{\infty} \lambda_i Z_i^2$, where the λ_i are solutions to an eigenvalue problem associated with the "kernel" $Q(x,y)$, and the Z_i are i.i.d. standard normal variables. Goodness of fit statistics often satisfy (1.2) (see Filippova (1962) and Gregory (1977) for examples). Extension to the case that the first nonvanishing derivative is the third or higher is straightforward in principle.

The intent of this present investigation is to study "second order" properties of first order functionals. That is, for functionals satisfying (1.1) we seek to find the limiting distribution of

$$n(T(F_n) - T(F) - \frac{1}{n}\sum_{i=1}^{n} V(X_i)) . \tag{1.3}$$

Our approach will make use of the first and second Frechet-type differentials (derivatives) $T(F;\Delta)$ and $T^2(F;\Delta_1,\Delta_2)$. In particular, the existence of such differentials will imply, under suitable restrictions, that

$$n(T(F_n) - T(F) - \frac{1}{n}\sum_{i=1}^{n} V(X_i) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q(X_i,X_j)) \xrightarrow{p} 0 , \quad n\to\infty . \tag{1.4}$$

The limiting distribution of (1.3) is then available from existing results regarding quadratic forms $n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q(X_i,X_j)$.

In Section 2 we present the relevant definitions and theorems to implement the approach. In Section 3 we briefly examine the connections between our results and those of C. R. Rao's on first and second order efficiency (e.g., Rao (1963)). In Section 4 we will establish (1.4) for L- and M-estimators. In particular, we compare the "Huber" M-estimators and the corresponding α-trimmed means with the same asymptotic variance.

2. Differentials. Let T be a real-valued functional defined on a convex set $\mathcal{J}$ of d.f.'s. Denote by $\mathcal{D}(\mathcal{J})$ the linear space generated by differences

H-G of members of $\mathcal{J}$, $\mathcal{D}(\mathcal{J}) = \{\Delta: \Delta=c(H-G), H,G\in\mathcal{J}, \text{ c real}\}$. Assume that $\mathcal{D}(\mathcal{J})$ is equipped with a norm $||\cdot||$.

DEFINITION. T has a <u>differential</u> at $F\in\mathcal{J}$ with respect to the norm $||\cdot||$ and the set $\mathcal{S}_F$ if there exists a quantity $T(F;\Delta)$ defined on $\Delta\in\mathcal{D}(\mathcal{J})$ and linear in Δ which satisfies

$$\lim_{\substack{||G-F||\to 0 \\ G\in\mathcal{S}_F}} \frac{T(G) - T(F) - T(F;G-F)}{||G-F||} = 0 \;. \;\square \tag{2.1}$$

$T(F;\Delta)$ is called the "differential." Relaxed versions of (2.1) were introduced in Boos and Serfling (1977) to make verification easier. In particular, if $||F_n-F|| \xrightarrow{p} 0$ and

$$\frac{T(F_n) - T(F) - T(F;F_n-F)}{||F_n-F||} \xrightarrow{p} 0 \;,\; n\to\infty \;, \tag{2.2}$$

then we say that T has a weak stochastic differential at F with respect to $||\cdot||$ and the sequence $\{X_i\}$ which generates F_n. The value of (2.2) can be seen from the following lemma whose proof is trivial. Let $\mu(T,F) = E_F\{T(F;\delta_X-F)\}$ and $\mu^2(T,F) = \mathrm{Var}\{T(F;\delta_X-F)\}$.

LEMMA 1. <u>Let $\{X_i\}$ be a sequence of r.v.'s (not necessarily independent) with distribution F. Suppose that T has a weak stochastic differential at F with respect to $||\cdot||$ and $\{X_i\}$. Suppose further that</u>

$$\sqrt{n}\,||F_n - F|| = O_p(1) \;,\; n\to\infty \;. \tag{2.3}$$

<u>Then</u>

$$\sqrt{n}\,(T(F_n) - T(F) - T(F;F_n-F)) \xrightarrow{p} 0 \;,\; n\to\infty \;. \tag{2.4}$$

<u>Moreover, if the X_i's are independent and</u> $\mu(T,F) = 0$ <u>and</u> $0 < \sigma^2(T,F) < \infty$, <u>then</u>

(2.5) $$\sqrt{n}\ (T(F_n) - T(F)) \xrightarrow{d} N(0,\sigma^2(T,F)),\ n\to\infty.\ \square$$

EXAMPLE. Let $T(F) = 1/\int x dF(x)$. Suppose F has mean $\mu \neq 0$ and variance σ^2. Then $T(F;\Delta) = -\int x d\Delta/\mu^2$, $T(F;F_n-F) = n^{-1}\sum_{i=1}^{n}(\mu-X_i)/\mu^2$, $\sqrt{n}\ (T(F_n) - T(F)) = \sqrt{n}\ (\bar{X}^{-1}-\mu^{-1}) \xrightarrow{d} N(0,\sigma^2/\mu^4)$.

Analogous to the definition of the (first) differential, we have

DEFINITION. T has a second differential at $F\epsilon\mathcal{J}$ with respect to the norm $||\cdot||$, the set $\mathcal{S}_F$, and the functional $T(F;\Delta)$, if there exists a quantity $T^2(F;\Delta_1,\Delta_2)$ defined on $(\Delta_1,\Delta_2)\epsilon\mathcal{D}(\mathcal{J})\times\mathcal{D}(\mathcal{J})$ and bilinear in (Δ_1,Δ_2) which satisfies

(2.6) $$\lim_{\substack{||G-F||\to 0 \\ G\epsilon\mathcal{S}_F}} \frac{T(G) - T(F) - T(F;G-F) - \frac{1}{2}T^2(F;G-F,G-F)}{(||G-F||)^2} = 0\ .\ \square$$

$T^2(F;\Delta_1,\Delta_2)$ is called the "second differential." Note that we do not assume anything about $T(F;\Delta)$ ($T(F;\Delta) = 0$ is allowed), though we implicitly expect (2.1) or (2.2) to hold. The analogous definition of a weak stochastic second differential follows for $G=F_n$.

A simple lemma will show how to find the candidate for $T^2(F;\Delta_1,\Delta_2)$ in most problems. Let $D_G T(F)$ be the usual right-hand derivative of $T(F+t(G-F))$ with respect to t evaluated at t=0. In Boos ((1977a), Lemma 3.1) it was shown that if T has a (first) differential at F with respect to $||\cdot||$ and $\mathcal{S}_F$, and $F_t = F+t(G-F)\in\mathcal{S}_F$ for all sufficiently small t, then $D_G T(F)$ exists and equals $T(F;G-F)$. The following lemma provides the analogous result for second differentials. Let $D_G^2 T(F)$ be the second right-hand derivative of $T(F_t)$ evaluated at t=0.

LEMMA 2. <u>Suppose that T has a first and second differential at F with respect to $||\cdot||$ and $\mathcal{S}_F$. Let $G\epsilon\mathcal{J}$ be given such that $F_t\epsilon\mathcal{S}_F$ for all sufficiently small t. Suppose that $D_G^2 T(F)$ exists. Then $D_G^2 T(F) = T^2(F;G-F,G-F)$.</u>

PROOF. Note that $F_t - F = t(G-F)$, so that $||F_t - F|| = t||G-F|| \to 0$ as $t \to 0+$ for fixed G. For small t $T(F;F_t-F) = tT(F;G-F) = tD_G T(F)$ by Lemma 3.1 of Boos (1977a). The definition of the second differential,

$$T(F_t) - T(F) - T(F;F_t-F) = \tfrac{1}{2}T^2(F;F_t-F,F_t-F) + o(||F_t-F||^2),\ t\to\infty ,$$

can then be rewritten as

$$(2.7) \qquad T(F_t) - T(F) - tD_G T(F) = t^2[\tfrac{1}{2}T^2(F;G-F,G-F) + o(1)] \ ,\ t\to 0+ \ .$$

By Taylor expansion of $T(F_t)$ about t=0, we have

$$(2.8) \qquad T(F_t) - T(F) - tD_G T(F) = \frac{t^2}{2} D_G^2 T(F) + o(t^2) \ ,\ t\to 0+ \ .$$

Combining (2.7) and (2.8), we get $D_G^2 T(F) = T^2(F;G-F,G-F)$. □

EXAMPLE. Let $T(F) = 1/\int x dF$ as in the previous example. Then

$$\frac{d}{dt} T(F_t) = \frac{-\int x d(G-F)}{[\int x dF_t]^2}$$

and

$$\frac{d^2}{dt^2} T(F_t) = \frac{2[\int x d(G-F)]^2}{[x dF_t]^3}$$

Setting t=0, we find $D_G T(F) = -\int x d(G-F)/\mu^2$ and $D_G^2 T(F) = 2[\int x d(G-F)]^2/\mu^3$. For a sequence $\{X_i\}$ of i.i.d. r.v.'s with distribution F such that $0 < \int |x| dF < \infty$, it is not difficult to show that $T(F;\Delta) = \int x d\Delta/\mu^2$ and $T^2(F;\Delta_1,\Delta_2) = 2\int x d\Delta_1 \int x d\Delta_2/\mu^3$ are weak stochastic first and second differentials with respect to $||\cdot||_\infty$ and $\{X_i\}$.

The usefulness of $T(F;\Delta)$ in obtaining asymptotic normality of $T(F_n)$ for first order functionals stems from the representation of $T(F;F_n-F)$ as an average $n^{-1}\sum_{i=1}^{n}V(X_i)$. Likewise, the usefulness of $T^2(F;\Delta_1,\Delta_2)$ stems from the representation of $T^2(F;F_n-F,F_n-F)$ as a simple quadratic form $n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}2Q(X_i,X_j)$. Both representations follow from linearity and the simple form of the empirical d.f. $F_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i}$:

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

$$= \frac{1}{n}\sum_{i=1}^{n}V(X_i)\ ,$$

$$T^2(F;F_n-F,F_n-F) = T^2(F;\frac{1}{n}\sum_{i=1}^{n}(\delta_{X_i}-F),\ \frac{1}{n}\sum_{j=1}^{n}(\delta_{X_i}-F))$$

$$= \frac{1}{n}\sum_{i=1}^{n}T^2(F;\delta_{X_i}-F\ \ ,\ \frac{1}{n}\sum_{j=1}^{n}(\delta_{X_j}-F))$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\ \sum_{j=1}^{n}T^2(F;\delta_{X_i}-F,\delta_{X_j}-F)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\ \sum_{j=1}^{n}2Q(X_i,X_j)\ ,$$

where for simplicity we have set $T(F;\delta_x-F) = V(x)$ and $T^2(F;\delta_x-F,\delta_y-F) = 2Q(x,y)$.

The asymptotic properties of such quadratic forms have been investigated by numerous authors including Filippova (1962) and Gregory (1977). The latter gives a general result for contiguous sequences. We reformulate his result for the simple i.i.d. situation. Let $Q(x,y)$ be a <u>symmetric</u>, nonzero kernel on R_2 such that

(2.9) $$\int Q^2(x,y)dF(x)dF(y) < \infty$$

and

(2.10) $$\int Q(\cdot,y)dF(y) = 0 \quad \text{a.e. } [F]$$

for a fixed d.f. F. Let $\{\lambda_k, k \geq 0\}$ denote the finite or infinite collection of eigenvalues of Q which satisfy $\int Q(x,y)f_k(y)dF(y) = \lambda_k f_k(x)$ a.e. [F], $\int f_k(x)f_j(x)dF(x) = 0$ if $k \neq j$, and $\int f_k^2(x)dF(x) = 1$. Let f_0 correspond to $\lambda_0=0$. The next lemma follows from Theorem 2.3 of Gregory (1977).

LEMMA 3. *Let* $\{X_i\}$ *be a sequence of independent r.v.'s with distribution* *F. Suppose that* $\sum_{k=1}^{\infty}\lambda_k < \infty$. *Then*

(2.11) $$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}Q(X_i,X_j) \xrightarrow{d} \sum_{k=1}^{\infty}\lambda_k Z_k^2, \quad n\to\infty,$$

where $Z_1, Z_2, \ldots$ *are i.i.d. standard normal variables.*

In light of (2.9) and (2.10) and the assumed independence of the X_i's, the mean and variance of $n^{-1}\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}Q(X_i,X_j)$ are given by $E_F Q(X,X)$ and $n^{-1}\text{Var}_F Q(X,X) + 2(1-n^{-1})E_F Q^2(X_1,X_2)$.

Let $L_2(R_1,F)$ be the function space consisting of all real-valued functions g such that $\int g^2 dF < \infty$. The eigenvalues defined above pertain to the Fredholm operator $Ag(x) = \int Q(x,y)g(y)dF(y)$ which maps $L_2(R_1,F)$ into itself. If $Q(x,y)$ is symmetric and (2.9) holds, then these operators are called Hilbert-Schmid operators. The theory of such operators is well-developed (see [22], Chs. 2 and 3, for a survey of results) and yields the representation $Q(x,y) = \Sigma_{k=1}^{\infty}\lambda_k f_k(x)f_k(y)$. This latter representation leads to $\int Q^2(x,y)dF(x)dF(y) = \sum_{k=1}^{\infty}\lambda_k^2$, and if further, $\sum_{k=1}^{\infty}|\lambda_k| < \infty$, then

$\int Q(x,x)dF(x) = \sum_{k=1}^{\infty}\lambda_k$. Of special interest are "degenerate" kernels having the form $Q(x,y) = \sum_{k=1}^{\ell} g_k(x)h_k(y)$ for functions $g_k, h_k \in L_2(R_1,F)$, $k=1,\ell$. The operators defined by such kernels have only a finite number of eigenfunctions and eigenvalues. In Section 4 we will show that the kernels associated with certain M-estimators have this simple form.

Combining Lemma 3 with our differential theory leads to the following theorem.

THEOREM 1. *Let $\{X_i\}$ be a sequence of r.v.'s (not necessarily independent) with distribution* F. *Suppose that* T *has a weak stochastic second differential with respect to* $||\cdot||$, $\{X_i\}$, and $T(F;\Delta)$. *Suppose further that* (2.3) *holds. Then*

$$n(T(F_n) - T(F) - T(F;F_n-F) - \tfrac{1}{2}T^2(F;F_n-F,F_n-F)) \xrightarrow{p} 0 \ , \ n\to\infty. \tag{2.12}$$

Moreover, if the X_i *are independent and* $Q(x,y) = \frac{1}{2}T^2(F;\delta_x-F,\delta_y-F)$ *is symmetric and satisfies* (2.9) *and* (2.10) *and* $\sum_{k=1}^{\infty}\lambda_k < \infty$, *then*

$$n(T(F_n) - T(F) - T(F;F_n-F)) \xrightarrow{d} \sum_{k=1}^{\infty}\lambda_k Z_k^2 \ , \ n\to\infty \ , \tag{2.13}$$

where $Z_1, Z_2, \ldots$ *are i.i.d. standard normal variables.*

PROOF. (2.12) follows trivially from (2.3) and the definition of weak stochastic second differential. Then (2.13) follows from (2.12) and Lemma 3. □

REMARKS. (i) If $T(F;F_n-F) \equiv 0$, then (2.13) just expresses the usual convergence of second order functionals. In this case $Q(x,y)$ might be interpreted as an "influence curve" for T. (ii) If (2.4) and (2.5) hold, then (2.13) expresses the rate at which $T(F_n) - T(F)$ is approximated by $T(F;F_n-F)$. We see two potential applications of this latter case. For estimators which are first order efficient as defined by Rao (1963), we suggest

$2E_F Q^2 = 2\int Q^2(x,y)dF(x)dF(y)$ as a measure of second order efficiency (the smaller the better). We compare this idea with Rao's definition of second order efficiency in the next section. For two estimators (or test statistics) whose approximating sums $T(F;F_n-F)$ are equal, though neither is first order efficient - as in the case of many robust estimators, we suggest $2E_F Q^2$ as one criterion to use in choosing between the two. In Section 4 we will compare L- and M-estimators on this basis.

EXAMPLE. Continuing our previous example, we find that $T^2(F;\Delta_1,\Delta_2) = 2\int x d\Delta_1 \int x d\Delta_2/\mu^3$ yields $Q(x,y) = (x-\mu)(y-\mu)/\mu^3$, and $\int Q(x,y)f(y)dF(y) = \lambda f(x)$ is satisfied by only functions of the form $f(x) = c(x-\mu)$. Normalizing, we get $f(x) = (x-\mu)/\sigma$ and $\lambda = \sigma^2/\mu^3$. Hence

$$n\left(\frac{1}{\bar{X}} - \frac{1}{\mu} - \frac{(\mu-\bar{X})}{\mu^2}\right) \xrightarrow{d} \frac{\sigma^2}{\mu^3} Z^2 \ , \ n\to\infty \ .$$

Note that $2E_F Q^2 = 2\sigma^4\mu^{-6} = \mathrm{Var}(\sigma^2\mu^{-3}Z^2)$.

3. Second Order Efficiency. In a series of papers in the early sixties (Rao (1960), (1961), (1962), (1963)), C. R. Rao developed a theory of asymptotic efficiency which is intimately connected with the likelihood function and Fisher information. More recent papers on the subject are Ghosh and Subramanyan (1974) and Efron (1975). Guiding themes in Rao's development were:

(a) "Efficient" estimators should in some sense summarize data without loss of information.

(b) Superefficiency should not occur.

(c) There should be a way to choose between efficient estimators.

Rather than attempt to assess Rao's contribution, we will give the essential definitions and show some relationships between his theory and that of Section 2.

Let $L(\underline{X};\theta)$ be the likehood function of a sample $\underline{X} = (X_1, \ldots, X_n)$, each X_i having d.f. F_θ and density $f(x;\theta)$, where θ is real-valued. For independent X_i, $L(\underline{X};\theta) = \pi_{i=1}^{n} f(X_i;\theta)$. In accord with (a) and (b) above, Rao (1961) defines $T_n = T_n(X_i, \ldots, X_n)$ to be "first order efficient" for estimation of θ if

$$(3.1) \qquad \sqrt{n}\left(\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\underline{X};\theta) - \frac{\alpha}{\sqrt{n}} - \beta(T_n-\theta)\right) \xrightarrow{p} 0, \; n\to\infty ,$$

where α and β are constants possibly depending on θ.

If the X_i are independent and Fisher information is finite, $\mathcal{J}(\theta) = E_{F_\theta}[\partial/\partial\theta \log f(X;\theta)]^2 < \infty$, and $\alpha = 0$, $\beta = \mathcal{J}(\theta)$, then (3.1) leads to the often-used asymptotic variance criterion of efficiency

$$(3.2) \qquad \sqrt{n}\,(T_n-\theta) \xrightarrow{d} N(0, \frac{1}{\mathcal{J}(\theta)}) , \; n\to\infty .$$

Of course, (3.1) is stronger than (3.2) and eliminates the usual superefficient pathologies.

The question posed by (c) seems to have originated in a multinomial model where a number of methods produced estimators satisfying (3.1). These methods included maximum likelihood and a number of minimum distance methods. In a motivating discussion, Rao ((1961), p. 538) first suggested the asymptotic variance of

$$(3.3) \qquad n\left(\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\underline{X};\theta) - \frac{\alpha}{\sqrt{n}} - \beta(T_n-\theta)\right)$$

as a measure of the rate of convergence in (3.2). However, he decided to use instead the asymptotic variance of

$$(3.4) \qquad n\left(\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\underline{X};\theta) - \frac{\alpha}{\sqrt{n}} - \beta(T_n-\theta) - \lambda(T_n-\theta)^2\right) ,$$

minimized with respect to λ. The minimum asymptotic variance of (3.4), say E_2, is called the second order efficiency of T_n. The motivation for this measure seems to be that under certain regularity conditions $E_2 = \lim_{n\to\infty}(ni - ni_T)$, where ni and ni_T are the information contained in the sample and in the statistic respectively. Thus E_2 is not so much a measure of the rate of convergence of the estimator T_n, but refers instead to some intrinsic property of T_n as a substitute for the whole sample (see Rao (1963), p. 200).

The following theorem shows how the asymptotic variance of (3.3) relates to E_2 in the case $\alpha = 0$, $\beta = \mathcal{I}(\theta)$. Let $U_n(\underline{X};\theta) = [n\mathcal{I}(\theta)]^{-1}\partial/\partial\theta \log L(\underline{X};\theta)$ and $Y_n = (T_n-\theta) + C(T_n-\theta)^2$, $C \in (-\infty,\infty)$.

THEOREM 2. Let $\{X_i\}$ be a sequence of independent r.v.'s with distribution F_θ. Suppose that $\mathcal{I}(\theta) < \infty$ and $E_{F_\theta}\{\partial/\partial\theta \log f(X;\theta)\} = 0$ and

(3.5) $$\sqrt{n}\,(T_n-\theta-U_n(\underline{X};\theta)) \xrightarrow{p} 0 \ , \ n\to\infty;$$

(3.6) $$n(T_n-\theta-U_n(\underline{X};\theta) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q(X_i,X_j)) \xrightarrow{p} 0 \ , \ n\to\infty \ .$$

Then

(3.7) $$\sqrt{n}(\,Y_n-U_n(\underline{X};\theta)) \xrightarrow{p} 0 \ , \ n\to\infty \ ,$$

and

(3.8) $$n(Y_n-U_n(\underline{X};\theta) - (\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q(X_i,X_j) + C[U_n(\underline{X};\theta)]^2)) \xrightarrow{p} 0, \ n\to\infty \ .$$

Note that (3.5) and (3.6) are conclusions of the form (2.4) and (2.12). Here $U_n(\underline{X};\theta)$ plays the role of $T(F_\theta;F_n-F_\theta)$. If $Q(x,y)$ is symmetric and satisfies (2.9) and (2.10) and $\Sigma_{k=1}^{\infty}\lambda_k < \infty$, then (3.6) and Lemma 3 tell us that the asymptotic variance of $n(T_n-\theta-U_n(\underline{X};\theta))$ is $\mathrm{Var}\,(\Sigma_{k=1}^{\infty}\lambda_k Z_k^2) = 2\Sigma_{k=1}^{\infty}\lambda_k^2 = 2E_{F_\theta}Q^2$. Likewise, if the same assumptions apply to the kernel

$$Q^*(x,y) = Q(x,y) + C(\frac{\partial}{\partial\theta}\log f(x;\theta))(\frac{\partial}{\partial\theta}\log f(y;\theta)),$$

then (3.8) and Lemma 3 tell us that the asymptotic variance of $n(Y_n - U_n(\underline{X};\theta))$ is $2E_{F_\theta} Q^{*2}$ and $E_2 = \min_c E_{F_\theta} Q^{*2}$. Thus E_2 is given by

$$E_2 = 2\left[\int Q^2(x,y)dF_\theta(x)dF_\theta(y) - \frac{\left(\int Q(x,y)\frac{\partial}{\partial\theta}\log f(x;\theta)\frac{\partial}{\partial\theta}\log f(y;\theta)dF_\theta(x)dF_\theta(y)\right)^2}{[\mathcal{J}(\theta)]^2}\right]$$

PROOF OF THEOREM 2. The basic assumptions and (3.5) imply that $\sqrt{n}(T_n-\theta)$ and $\sqrt{n}U_n(\underline{X};\theta) = n^{-\frac{1}{2}}\Sigma_{i=1}^n[\mathcal{J}(\theta)]^{-1}\partial/\partial\theta \log f(X_i;\theta)$ each converge in distribution to $N(0,[\mathcal{J}(\theta)]^{-1})$. Then

$$\begin{aligned}\sqrt{n}\,(Y_n - U_n(\underline{X};\theta)) &= \sqrt{n}\,(T_n-\theta-U_n(\underline{X};\theta)) \\ &\quad + \sqrt{n}\,C(T_n-\theta)^2 .\end{aligned} \tag{3.9}$$

The first term on the right-hand side of (3.9) $\xrightarrow{p}$ 0 by (3.5), and the second term $\xrightarrow{p}$ 0 by the asymptotic normality of T_n. For (3.8) it is enough to show

$$n((T_n-\theta)^2 - (U_n(\underline{X};\theta))^2) \xrightarrow{p} 0 . \tag{3.10}$$

Factoring (3.10), we have

$$\sqrt{n}(T_n-\theta + U_n(\underline{X};\theta))\cdot\sqrt{n}(T_n-\theta-U_n(\underline{X};\theta)).$$

The first factor is $O_p(1)$ by the asymptotic normality of $\sqrt{n}(T_n-\theta)$ and $\sqrt{n}U_n(\underline{X};\theta)$. The second factor is $o_p(1)$ by (3.5). □

EXAMPLE. Let $T(G) = [\int x dG]^2$. Suppose that $F_\theta(x) = \Phi(x-\sqrt{\theta})$, where Φ is the standard normal. Then $T(F_\theta) = \theta$, $\mathcal{I}(\theta) = 1/4\theta$ for $\theta \neq 0$, $\partial/\partial\theta \log f(x;\theta) = (x-\sqrt{\theta})/2\sqrt{\theta}$, $U_n(\underline{X};\theta) = (2\sqrt{\theta}\, n)^{-1}\sum_{i=1}^n (X_i-\sqrt{\theta})$, $Q(x,y) = (x-\sqrt{\theta})(y-\sqrt{\theta})$, and $Q^*(x,y) = Q(x,y)[1 + C\mathcal{I}(\theta)]$. Thus $2E_{F_\theta} Q^2 = 2$ and $E_2 = \min_C E_{F_\theta} Q^{*2} = 0$.

4. L- and M-Statistics.

4.1 Introduction. Let $F^{-1}(t) = \inf\{x:F(x) \geq t\}$. Define the L-functional T_L by

$$T_L(F) = \int_0^1 F^{-1}(t)J(t)dt \ , \tag{4.1}$$

where $J(t)$ is a given "score" function. In Boos (1977b) conditions are given for T_L to have a differential

$$T_L(F;\Delta) = -\int \Delta(x)J(F(x))dx \tag{4.2}$$

and for

$$\sqrt{n}(T_L(F_n) - T_L(F)) \xrightarrow{d} N(0,\sigma^2(J,F)),\ n\to\infty \ , \tag{4.3}$$

where $\sigma^2(J,F) = E_F[\int(I(X \leq t) - F(t))J(F(t))dt]^2$. The defining equation (4.1) allows for both location functionals and scale functionals. For example, $J(t) = I(\alpha \leq t \leq 1-\alpha)/(1-2\alpha)$ yields a form of the α-trimmed mean, and $J(t) = t-1/2$ yields a form of Gini's mean difference.

The location M-functional T_M is defined to be the solution of

$$\lambda_F(c) = \int \psi(x-c)dF(x) = 0 \ , \tag{4.4}$$

where ψ is real-valued and often skew-symmetric about 0. If (4.4) has more than one solution, then some additional rule must be given to define T_M uniquely. In Boos and Serfling (1977) conditions are given for T_M to have a strong stochastic quasi-differential (another variant of the first differential) given by

$$(4.5) \qquad T_M(F;\Delta) = \frac{\int \psi(x - T_M(F))d\Delta(x)}{-\lambda'(T_M(F))}$$

and for

$$(4.6) \qquad \sqrt{n}\,(T_M(F_n) - T_M(F)) \xrightarrow{d} N(0, \sigma^2(\psi, F)),\ n \to \infty,$$

where $\sigma^2(\psi,F) = \int \psi^2(x - T_M(F))dF(x)/(\lambda'(T_M(F)))^2$. The defining equation (4.4) applies to simple location problems (scale known). If scale is unknown, then $\psi(x)$ can be replaced by $\psi(x/\hat{\sigma})$, where $\hat{\sigma}$ is some estimate of scale.

When F and ψ are suitably restricted, Jaeckel (1971) showed an interesting relationship between L- and M-functionals: setting $J = \psi'(F^{-1}(t))$ yields $\sigma^2(J,F) = \sigma^2(\psi,F)$. The differential approach shows even more: the first order approximating sums $T_L(F;F_n - F)$ and $T_M(F;F_n - F)$ are equal. More specifically, Lemma 1 and (4.2) and (4.5) lead to

$$(4.7) \qquad \sqrt{n}\,(T_L(F_n) - T_L(F) - T_L(F;F_n - F)) \xrightarrow{p} 0,\ n \to \infty,$$

and

$$(4.8) \qquad \sqrt{n}\,(T_M(F_n) - T_M(F) - T_M(F;F_n - F)) \xrightarrow{p} 0,\ n \to \infty,$$

where

$$(4.9) \qquad T_L(F;F_n-F) = \frac{1}{n}\sum_{i=1}^{n}\left(-\int[I(X_i\le t)-F(t)]J(F(t))dt\right)$$

and

$$(4.10) \qquad T_M(F;F_n-F) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\psi(X_i)-T_M(F)}{-\lambda'_F(T_M(F))}\right)$$

Let F be symmetric about 0, continuous, and strictly increasing. Let ψ be skew-symmetric about 0, i.e. $\psi(-x) = -\psi(x)$, and differentiable a.e. $[F]$ with $-\lambda'_F(0) = \int\psi'(x)dF(x) = 1$. Set $J(t) = \psi'(F^{-1}(t))$. Then $T_M(F) = T_L(F) = 0$, (4.10) becomes $n^{-1}\sum_{i=1}^{n}\psi(X_i)$, $J(F(t)) = \psi'(t)$, and (4.9) becomes $n^{-1}\sum_{i=1}^{n}(-\int[I(X_i\le t)-F(t)]\psi'(t)dt)$. Integration by parts yields term by term equality between (4.9) and (4.10),

$$\begin{aligned}-\int[I(X_i\le t)-F(t)]\psi'(t)dt &= \int\psi(t)d[I(X_i\le t)-F(t)]\\ &= \psi(X_i).\end{aligned}$$

4.2. <u>Second order theorems for L- and M-statistics</u>. In order to further compare the estimators $T_L(F_n)$ and $T_M(F_n)$, we need to look at their second order distributions. Theorem 3 establishes the existence of a second differential for T_L under mild restrictions on F and J. A corollary then gives the limiting distribution of $n(T_L(F_n) - T_L(F) - T_L(F;F_n-F))$. Theorem 4 provides directly the limiting distribution of $n(T_M(F_n) - T_M(F) - T_M(F;F_n-F))$ without proving the existence of a second differential.

Let $\mathcal{F}_J = \{F: |\int F^{-1}(t)J(t)dt| < \infty\}$ and $\mathcal{G}_F = \{G: S_G \subset S_F\}$, where S_G = support of G. For a bounded positive function q on $(0,1)$ we define the norm

(4.11) $$||\Delta||_{q(F)} = \sup_{x_1<x<x_2} \left|\frac{\Delta(x)}{q(F(x))}\right| , \quad \Delta\in\mathcal{D}(\mathcal{F}_J) ,$$

where $[x_1,x_2]$ is the smallest interval (possibly infinite) containing S_F.

THEOREM 3. Suppose that J is Lipschitz of order 1 on [0,1] and has a bounded derivative J' at all but a finite set B of F^{-1} measure zero. If

(4.12) $$\int q^2(F(x))\,dx < \infty ,$$

then T_L has a second differential at $F\in\mathcal{F}_J$ with respect to $||\cdot||_{q(F)}$ and $\mathcal{S}_F$ given by

(4.13) $$T^2(F;\Delta_1,\Delta_2) = -\int\Delta_1(x)\Delta_2(x)J'(F(x))dx.$$

PROOF. Since (4.13) is clearly bilinear in (Δ_1,Δ_2), we need to show

(4.14) $$T(G)-T(F) - \int-(G-F)J(F)dx - \tfrac{1}{2}\int-(G-F)^2J'(F)dx = o(||G-F||^2_{q(F)}),\ ||G-F||_{q(F)}\to 0.$$

(Inside the integrals we suppress the fact that F and G depend on x.) In Boos (1977b) it is shown that $T(G) - T(F) = \int(K(F)-K(G))dx$, where $K(y) = \int_0^y J(u)du$. Note that $K'(y) = J(y)$. Thus

$$|T(G) - T(F) - \int - (G-F)J(F)dx - \tfrac{1}{2}\int - (G-F)^2J'(F)dx|$$

$$= |\int_{R_1-B} [K(G) - K(F) - (G-F)J(F) - \tfrac{1}{2}(G-F)^2J'(F)]dx|$$

(4.15) $$\le ||G-F||^2_{q(F)}\int_{R_1-B} q^2(F(x))W(x)dx ,$$

where

$$W(x) = \frac{K(G(x))-K(F(x))-[G(x)-F(x)]J(F(x))-\frac{1}{2}[G(x)-F(x)]^2 J'(F(x))}{[G(x) - F(x)]^2}, \quad G(x) \neq F(x),$$

$$= 0, \quad G(x) = F(x).$$

Note that for each $x \in R_1 - B$ we have $\lim_{||G-F||_{q(F)} \to 0} W(x) = 0$ by simple Taylor expansions (see Cartán (1971), Theorem 5.6.3). To show that (4.14) follows from (4.15), it is only necessary to show that lim and $\int$ can be interchanged in (4.15). This interchange is allowed via dominated convergence by (4.12) and a **bound** on $W(x)$: for $G(x) \neq F(x)$

$$|W(x)| = |K(G(x))-K(F(x))-[G(x)-F(x)]J(F(x))-\tfrac{1}{2}[G(x)-F(x)]^2 J'(F(x))|/[G(x)-F(x)]^2$$

$$= \Big|\int_{F(x)}^{G(x)} [J(u)-J(F(x))]\,du - \tfrac{1}{2}[G(x)-F(x)]^2 J'(F(x))\Big|/[G(x)-F(x)]^2$$

$$\leq \Big|\int_{F(x)}^{G(x)} A|u-F(x)|du\Big|/[G(x)-F(x)]^2 + \tfrac{1}{2}||J'||_\infty$$

$$\leq A + \tfrac{1}{2}||J'||_\infty. \ \square$$

REMARKS. (i) If J is trimmed in neighborhoods of 0 and 1, then (4.12) is unnecessary and the conclusion of Theorem 3 holds for $||\cdot||_{q(F)}$ replaced by $||\cdot||_\infty$. (ii) A popular choice of q is

(4.16) $$q(x) = [x(1-x)]^{\frac{1}{2}-\delta}, \quad 0 < \delta < \tfrac{1}{2}.$$

In Boos (1977b) it is shown that

$$(4.17) \qquad \sqrt{n}||F_n - F||_{q(F)} = O_p(1) \;,\; n\to\infty \;,$$

for q in a certain class (including (4.16)) and for F_n generated by an i.i.d sequence from F (F arbitrary). (iii) We reemphasize the lack of restrictions on F, such as continuity, in the statement of Theorem 3.

Remark (ii) and Theorem 1 provide an application of Theorem 3 in the following corollary, whose proof should be transparent.

COROLLARY. <u>Suppose that the conditions of Theorem 3 hold with q given by (4.16). Let $\{X_i\}$ be a sequence of i.i.d. r.v.'s with distribution F. Then</u>

$$(4.18) \qquad n[T_L(F_n) - T_L(F) - T_L(F;F_n-F) - \tfrac{1}{2}\int - (F_n-F)^2 J'(F)dx] \xrightarrow{p} 0,\; n\to\infty.$$

<u>Moreover, if $Q_L(x,y) = -\frac{1}{2}\int[I(x\le t)-F(t)][I(y\le t)-F(t)]J'(F(t))dt$ has eigenvalues λ_k such that $\sum_{k=1}^{\infty}\lambda_k < \infty$, then</u>

$$(4.19) \qquad n(T_L(F_n) - T_L(F) - T_L(F;F_n-F)) \xrightarrow{d} \sum_{k=1}^{\infty}\lambda_k Z_k^2 \;,\; n\to\infty,$$

<u>where $Z_1, Z_2, \ldots$ are i.i.d. standard normal variables.</u>

EXAMPLES. (i) The trimmed mean, $J(u) = I(\alpha\le u\le 1-\alpha)/(1-2\alpha)$, is not Lipschitz. However, a smoothly trimmed version can be defined as follows. For $\epsilon > 0$ let $J(u) = 0$ for $u\le\alpha-\epsilon$, let J linearly slope up to $[\epsilon + (1-2\alpha)]^{-1}$ at $u = \alpha$, let $J(u) = [\epsilon + (1-2\alpha)]^{-1}$ for $\alpha\le u\le 1-\alpha$, and let J linearly slope back to zero at $u = 1-\alpha+\epsilon$ and remain zero for $u > 1-\alpha+\epsilon$. Then J is Lipschitz (of order 1) on [0,1] and has a derivative except at $\alpha-\epsilon$, α, $1-\alpha$, and $1-\alpha+\epsilon$. Remark (i) applies here, so that (4.12) is unnecessary and the other hypotheses of Theorem 3 are satisfied for any F with unique quantiles at $F^{-1}(\alpha-\epsilon)$, $F^{-1}(\alpha)$,

$F^{-1}(1-\alpha)$, and $F^{-1}(1-\alpha+\varepsilon)$. Note that as $\varepsilon\to 0$, this functional approaches the trimmed mean, and we anticipate the validity of Theorem 3 under relaxed conditions on J and strengthened conditions on F. (ii) Gini's mean difference, $J(u) = u-\frac{1}{2}$.

Turning to M-statistics, we establish directly the analogues of (4.18) and (4.19). The origin of the M-estimator T_n is left unspecified, subject only to (4.24) and (4.25), in order to allow flexibility of definition. See Boos and Serfling (1977) and Collins (1976) for two different methods of Choosing T_n. Let $||\cdot||_V$ be the usual variation norm

$$||g||_V = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \sup \sum_{i=0}^{n} |g(x_{i+1}) - g(x_i)|,$$

the sup being taken over all partitions $a = x_0 < x_1 <\dots< x_n < x_{n+1} = b$ of $[a,b]$.

THEOREM 4. _Let_ F _and_ ψ _be such that_ $\lambda_F(T_o) = 0$, $\lambda_F'(T_o) \neq 0$, _and_ $\lambda_F''(T_o) = 0$. _Assume that_ ψ' _exists everywhere and is continuous_. _Let_ ψ _and_ ψ' _satisfy_

(4.20) $$\lim_{b\to 0} ||\psi(x+b) - \psi(x)||_V = 0 ;$$

(4.21) $$\lim_{b\to 0} ||\psi'(x+b) - \psi'(x)||_V = 0 ;$$

(4.22) $$\int\psi^2(x-T_o)dF(x) < \infty ;$$

(4.23) $$\int[\psi'(x-T_o)]^2dF(x) < \infty .$$

Let $\{X_i\}$ be a sequence of i.i.d. r.v.'s with distribution F. *Let the sequence* $T_n = T_n(X_1,\ldots X_n)$ *satisfy*

(4.24) $$P(\lambda_{F_n}(T_n)=0) \to 1, \ n\to\infty ,$$

and

(4.25) $$T_n \xrightarrow{p} T_o , \ n\to\infty .$$

Then

(4.26) $$n(T_n-T_o-T_M(F;F_n-F) - \tfrac{1}{2}T_M^2(F;F_n-F,F_n-F)) \xrightarrow{p} 0, \ n\to\infty ,$$

where $T_M(F;\Delta) = \int\psi(x-T_o)d\Delta(x)/(-\lambda_F'(T_o))$ *and*

(4.27) $$T_M^2(F;\Delta_1,\Delta_2) = \frac{-2\int\psi(x-T_o)d\Delta_1(x)\int\psi'(x-T_o)d\Delta_2(x)}{(\lambda_F'(T_o))^2}$$

If, further, $Q_M(x,y) = \psi(x-T_o)(\int\psi'(x-T_o)dF(x) - \psi'(y))/(\lambda_F'(T_o))^2$ *has eigenvalues* λ_k *such that* $\Sigma_{k=1}^{\infty}\lambda_k < \infty$, *then*

(4.28) $$n(T_n-T_o-T_M(F;F_n-F)) \xrightarrow{d} \sum_{k=1}^{\infty}\lambda_k Z_k^2 , \ n\to\infty ,$$

where $Z_1,Z_2,\ldots$ *are i.i.d. standard normal variables*.

REMARKS. (i) If $\psi(x) = -\psi(-x)$ and F is symmetric about 0, then $T_o=0$ and $\lambda_F''(0) = 0$ (provided $\lambda_F''(0)$ exists). (ii) (4.22) and (4.23) are used to show that certain sums involving ψ and ψ' satisfy the central limit theorem. (iii) (4.20) and (4.21) are essential in our method of proof and arise from use of the following lemma found in Boos (1977a).

LEMMA 4. Let the function h be continuous and of bounded variation on $(-\infty,\infty)$. Let G and F be d.f.'s. Then

(4.29) $$\left|\int h\,d(G-F)\right| \leq ||h||_V ||G-F||_\infty .$$

PROOF OF THEOREM 2. By the mean value theorem

(4.30) $$\lambda_F(T_n) - \lambda_F(T_o) = (T_n - T_o)\lambda'_F(T_n^*) ,$$

where T_n^* lies between T_n and T_o (and hence $T_n^* \xrightarrow{p} T_o$, $n\to\infty$). Since $\lambda'_F(T_o) \neq 0$, we can choose N_1 large enough so that $P(\lambda'_F(T_n^*) \neq 0) > 1-\varepsilon\ \forall n>N_1$. Likewise we can choose N_2 large enough so that $P(\lambda_{F_n}(T_n) = 0) > 1-\varepsilon\ \forall n>N_2$. The following arguments are all conditional on the event $\{\lambda'_F(T_n^*) \neq 0\} \cap \{\lambda_{F_n}(T_n) = 0\}$ which has probability $> 1-\varepsilon\ \forall\ n>\max(N_1,N_2)$. From (4.30) we have

$$T_n - T_o = \frac{\lambda_F(T_n) - \lambda_F(T_o)}{\lambda'_F(T_n^*)} = \frac{\lambda_{F_n}(T_n) - \lambda_F(T_n)}{-\lambda'_F(T_n^*)} .$$

We can express $T_M(F;F_n-F)$ in a congruent manner by adding and subtracting a remainder term,

$$T_M(F;F_n-F) = \frac{\lambda_{F_n}(T_o) - \lambda_F(T_o)}{-\lambda'_F(T_n^*)} + [\lambda_{F_n}(T_o) - \lambda_F(T_o)]\left[\frac{1}{\lambda'_F(T_n^*)} - \frac{1}{\lambda'_F(T_o)}\right] .$$

Then

(4.31) $$T_n - T_o - T(F;F_n-F) = \frac{\phi_n(T_n) - \phi_n(T_o)}{-\lambda'_F(T_n^*)} + R_{1n} ,$$

where $\phi_n(t) = \lambda_{F_n}(t) - \lambda_F(t)$ and $R_{1n} = -\phi_n(T_o)[(\lambda'_F(T_n^*))^{-1} - (\lambda'_F(T_o))^{-1}]$.

At this point we show that (4.31) is $o_p(n^{-\frac{1}{2}})$. The first term on the right-hand side of (4.31) is bounded by $||F_n-F||_\infty||\psi(x-T_n) - \psi(x-T_o)||_V/|\lambda_F'(T_n^*)|$ which is $o_p(n^{-\frac{1}{2}})$ by (4.20) and (4.25) and the fact that $||F_n-F||_\infty$ is $O_p(n^{-\frac{1}{2}})$. The remainder term R_{1n} is $o_p(n^{-\frac{1}{2}})$ by the central limit theorem applied to $\phi_n(T_o) = n^{-1}\Sigma\psi(X_i-T_o)$ and the convergence $\lambda_F'(T_M^*) \xrightarrow{p} \lambda_F'(T_o)$.

The next step is to express $\frac{1}{2}T_M^2(F;F_n-F,F_n-F)$ in a suitable form:

$$\frac{1}{2}T_M^2(F;F_n-F,F_n-F) = T_M(F;F_n-F)\,\frac{\phi_n'(T_o)}{-\lambda_F'(T_o)}$$

$$= \frac{(T_n-T_o)}{-\lambda_F'(T_n^*)}\,\phi_n'(T_o) + R_{2n} + R_{3n}\ ,$$

where

$$R_{2n} = \left[T_n-T_o - T_M(F;F_n-F)\right]\frac{\phi_n'(T_o)}{\lambda_F'(T_o)}$$

and

$$R_{3n} = (T_n-T_o)\phi_n'(T_o)\left[\frac{1}{\lambda_F'(T_n^*)} - \frac{1}{\lambda_F'(T_o)}\right]\ .$$

Thus

$$|T_n - T_o - T_M(F;F_n-F) - \tfrac{1}{2}T_M^2(F;F_n-F,F_n-F)|$$

$$\le\left|\frac{1}{\lambda_F'(T_n^*)}\right|\ |\phi_n(T_n) - \phi_n(T_o) - (T_n-T_o)\phi_n'(T_o)|$$

$$+ |R_{1n}| + |R_{2n}| + |R_{3n}|\ .$$

It is sufficient to show that each of these 4 terms is $o_p(n^{-1})$. By the mean value theorem

$$|\phi_n(T_n) - \phi_n(T_o) - (T_n - T_o)\phi_n'(T_o)| = |T_n - T_o||\phi_n'(T_n^{**}) - \phi_n'(T_o)| ,$$

where T_n^{**} lies between T_n and T_o. Since $\sqrt{n}(T_n - T_o)$ is asymptotically normal (via (4.31) and the central limit theorem), we need to show $|\phi_n'(T_n^{**}) - \phi_n'(T_o)| = o_p(n^{-\frac{1}{2}})$. By Lemma 4

$$|\phi_n'(T_n^{**}) - \phi_n'(T_o)| \leq ||F_n - F||_\infty ||\psi'(x - T_n^{**}) - \psi'(x - T_o)||_V .$$

Condition (4.21) coupled with $T_n^{**} \xrightarrow{p} T_o$ yields the result. By similar arguments it is easy to show that R_{2n} and R_{3n} are $o_p(n^{-1})$. However, it is not clear that R_{1n} is $o_p(n^{-1})$. This will follow if $\sqrt{n}(\lambda_F'(T_n^*) - \lambda_F'(T_O)) \xrightarrow{p} 0$. Write

$$|\lambda_F'(T_n^*) - \lambda_F'(T_o)| = \frac{|\lambda_F'(T_n^*) - \lambda_F'(T_o)|}{|T_n^* - T_o|} \; |T^* - T_o| .$$

The first factor $\xrightarrow{p} \lambda_F''(T_o) = 0$ and the second factor is $O_p(n^{-\frac{1}{2}})$. □

EXAMPLE. $\psi(x) = \arctan x$. The solution T_n of $\lambda_{F_n}(c) = 0$ is unique since ψ is strictly increasing. Conditions (4.20) and (4.21) are easy to verify, and (4.25) follows in the i.i.d. situation from the results of Huber (1964). Any symmetric F will satisfy the other requirements of Theorem 4.

The "degenerate" form of the kernel Q_M lends itself easily to calculation of the associated eigenvalues. For the special case that $\psi(x) = -\psi(-x)$, F is symmetric about 0, and $\int\psi'(x)dF(x) = 1$, the kernel reduces to $Q_M(x,y) = \psi(x)(1-\psi'(y))$. Since our theory requires Q_M to be symmetric, we use $\tilde{Q}_M = [\psi(x)(1-\psi'(y)) + \psi(y)(1-\psi'(x))]/2$ in place of Q_M, noting that $\Sigma_{i=1}^n \Sigma_{j=1}^n Q_M(X_i, X_j) = \Sigma_{i=1}^n \Sigma_{j=1}^n \tilde{Q}_M(X_i, X_j)$.

LEMMA 5. *Let the d.f.* F *be symmetric about* 0. *Suppose that* $\psi(x) = -\psi(-x)$ *and that* $\int\psi'(x)dF(x) = 1$. *Then the eigenvalues and eigenvectors of* $\tilde{Q}_M$ *are given by*

$$\lambda_1 = \tfrac{1}{2}\sqrt{c_1c_2} \qquad \lambda_2 = -\tfrac{1}{2}\sqrt{c_1c_2} \tag{4.32}$$

and

$$f_1(x) = \frac{\psi(x)}{\sqrt{2c_1}} + \frac{1-\psi'(x)}{\sqrt{2c_2}} \qquad f_2(x) = \frac{\psi(x)}{\sqrt{2c_1}} - \frac{(1-\psi'(x))}{\sqrt{2c_2}}, \tag{4.33}$$

where $c_1 = \int\psi^2(x)dF(x)$ *and* $c_2 = \int(\psi'(x))^2dF(x) - 1$.

PROOF. All eigenvectors must be of the form $a\psi(x) + b(1-\psi'(x))$. Normalizing yields the constants. □

Under the conditions of Theorem 4 and Lemma 5, (4.28) becomes

$$n(T_n-T_o-T_M(F;F_n-F)) \xrightarrow{d} \lambda_1Z_1^2 + \lambda_2Z_2^2 . \tag{4.34}$$

Note that Var $(\lambda_1Z_1^2 + \lambda_2Z_2^2) = c_1c_2 = 2E_F\tilde{Q}_M^2$.

The L-statistic kernel Q_L is much harder to handle. For F uniform (0,1), De Wet and Venter (1973) show that the eigenvalues associated with Q_L satisfy a Sturm-Liouville-type equation, which simplifies calculations. In general we shall be satisfied with calculating the anticipated variance of the limiting second order distribution, i.e., $2E_FQ_L^2$, where

$$E_FQ_L^2 = \tfrac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F(s),F(t)) - F(s)F(t)]^2J'(F(s))J'(F(t))dsdt.$$

This last expression can be justified under the assumptions of Theorem 3 with q given by (4.16) via a version of Fubini's theorem.

In the special situation that $T_L(F) = T_o$ and $T_L(F;F_n-F) = T_M(F;F_n-F)$, results (4.18) and (4.26) allow calculation of the limiting distribution of $n(T_L(F_n) - T_n)$. Jaeckel (1971) has given the weaker result $T_L(F_n) - T_n = o_p(n^{-1})$.

THEOREM 5. Let the d.f. F be symmetric about 0, continuous, and strictly increasing. Let $\psi(x) = -\psi(-x)$, $\int\psi'(x)dF(x) = 1$, and set $J(t) = \psi'(F^{-1}(t))$. Suppose that $\{X_i\}$ is a sequence of independent r.v.'s with distribution F such that (4.18) and (4.26) hold. Then

$$(4.35)\qquad n(T_L(F_n)-T_n - (\int[Q_L(x,y)-Q_M(x,y)]dF_n(x)dF_n(y))) \xrightarrow{p} 0.$$

Moreover, if the eigenvalues λ_k of $Q_D = Q_L - \tilde{Q}_M$ satisfy $\Sigma_{k=1}^{\infty}\lambda_k < \infty$, then

$$(4.36)\qquad n(T_L(F_n)-T_n) \xrightarrow{d} \Sigma_{k=1}^{\infty}\lambda_k Z_k^2 ,$$

where $Z_1, Z_2, \ldots$ are i.i.d. standard normal variables.

PROOF. The conditions on F and ψ insure $T_L(F) = T_o$ and $T_L(F;F_n-F) = T_M(F;F_n-F)$. Subtraction of (4.26) from (4.18) gives (4.35). Application of Lemma 3 then gives (4.36). □

4.3. Hubers vs. α-trimmed means. The "Huber" family of M-estimators, $\psi(x) = \max(-k,\min(k,x))$, and the α-trimmed means, $J(t) = I(\alpha \leq t \leq 1-\alpha)/(1-2\alpha)$, are asymptotically equivalent to first order if $F(-k) = \alpha$ for any d.f. F symmetric about 0 with $f(-k) = f(k) > 0$. We would like to compare their second order asymptotic variances. Although Theorems 3 and 4 do not apply directly (since ψ' doesn't exist at -k and k and J is discontinuous at α and $1-\alpha$), we expect the results of these theorems to remain true for F sufficiently regular in neighborhoods of -k and k. In fact we can approximate ψ and J as closely as we want by ψ^* and J^* which do satisfy Theorems 3 and 4. For example, the linearly trimmed means mentioned previously provide such an approximation for the α-trimmed mean.

If J is of bounded variation and if F has density f such that $f[F^{-1}(t)]$ is positive on the support of J, Q_L may be expressed as

$$(4.37) \qquad Q_L(x,y) = -\tfrac{1}{2}\int_0^1 \frac{[I(x \le F^{-1}(t))-t][I(y \le F^{-1}(t))-t]}{f[F^{-1}(t)]}\, dJ(t) \quad .$$

For the α-trimmed mean (4.37) becomes

$$(4.38) \qquad Q_\alpha(x,y) = \frac{-1}{2(1-2\alpha)}\left[\frac{(I(x \le F^{-1}(\alpha))-\alpha)(I(y \le F^{-1}(\alpha))-\alpha)}{f[F^{-1}(\alpha)]} - \frac{(I(x \le F^{-1}(1-\alpha))-(1-\alpha))(I(y \le F^{-1}(1-\alpha))-(1-\alpha))}{f[F^{-1}(1-\alpha)]}\right]$$

Note that the kernel Q_α is symmetric and degenerate, allowing potentially easy calculation of eigenvalues. If $f[F^{-1}(\alpha)] = f[F^{-1}(1-\alpha)]$, the variance of the limiting second order distribution is given by

$$(4.39) \qquad 2E_F Q_\alpha^2 = 2\int Q_\alpha^2(x,y)dF(x)dF(y) = \frac{\alpha^2}{(1-2\alpha)(f[F^{-1}(\alpha)])^2} \ .$$

Similarly, letting $\psi_k(x) = \max(-k,,\min(k,x))/(1-2\alpha)$ so that $\int \psi_k'(x)dF(x) = 1$, we have $\widetilde{Q}_k(x,y) = [\psi_k(x)(1-\psi_k'(y)) + \psi_k(y)(1-\psi_k'(x))]/2$ and

$$(4.40) \qquad 2\ E_F\widetilde{Q}_k^2 = A_k(F)\left(\frac{2\alpha}{1-2\alpha}\right),$$

where we have written $A_k(F)$ for the first order asymptotic variance $E\psi_k^2$.

In Tables 1-4 we have computed (4.39) and (4.40) for a variety of situations. In almost all these situations the Huber outperforms the corresponding α-trimmed mean. However, in most real life problems, scale is unknown, thus necessitating a scale invariant version of the M-functional T_M. One such version is the solution $T_{M,S}$ of $\lambda_F(C/S(F)) = 0$, where S(F) is some scale functional. Boos (1977a) has given conditions for $T_{M,S}$ to have a strong stochastic quasi-differential $T_{M,S}(F;\Delta)$ when $S(F;\Delta)$ exists. In particular, if F is symmetric and $S(F) = 1$, then $T_{M,S}(F;\Delta)$ is given by (4.5) and $T_{M,S}(F_n)$ has essentially the same first order properties as $T_M(F_n)$. Although we have not proven the existence of a second differential for $T_{M,S}$, straightforward calculation of $D_G^2 T_{M,S}(F)$ allows us to anticipate the form of $T_{M,S}(F;\Delta_1,\Delta_2)$. Under the additional assumption that $S(F;\delta_x - F)$ is an even function about 0 and $\int\psi'(x)dF(x) = 1$, we expect

$$\tfrac{1}{2}T_{M,S}(F;\Delta_1,\Delta_2) = \tfrac{1}{2}T_M(F;\Delta_1,\Delta_2) + T(F;\Delta_1)S(F;\Delta_2)B - S(F;\Delta_1)\int\psi'(x)x d\Delta_2 , \tag{4.41}$$

where $B = 1 + \int\psi''(x)xdF(x)$. This leads to the unsymmetric kernel

$$Q_{M,S}(x,y) = \psi(x)(1-\psi'(y)) + \psi(x)S(F;\delta_y-F)B - S(F;\delta_x-F)\psi'(y)y \tag{4.42}$$

and second order asymptotic variance

$$2E_F\tilde{Q}^2_{M,S} = A_T(F)\left(\frac{2\alpha}{1-2\alpha}\right) + A_S(F)E[\psi(X)B-\psi'(X)X]^2 \tag{4.43}$$

$$+ 2E[S(F;\delta_X-F)(1-\psi'(X))][A_T(F)B-E\psi'(X)X],$$

where $A_T(F)$ and $A_S(F)$ are the first order asymptotic variances of $T_{M,S}(F_n)$ and $S(F_n)$. In Tables 1-4 we have included (4.43) for $\psi=\psi_k$ using $S(F) = (F^{-1}(3/4) - F^{-1}(1-4))/d_F$, a normalized interquantile range, where d_F is chosen for each F so that $S(F) = 1$. Now the α-trimmed mean appears very competitive. In the middle values, $.10 \le \alpha \le .25$, the α-trimmed mean tends to outperform the Huber scale version for the first three distributions. This same pattern persisted in a number of other distributions not displayed.

However, for $F = 0.75\,\Phi + .25\Phi(x/10)$ (not displayed) the Huber scale version was better than the α-trimmed mean for all values of k and α except k=1, α=.234. For the Cauchy (also not displayed), with density $f(x) = \pi^{-1}(1+x^2)^{-1}$, the α-trimmed mean wins at only 4 values of k and α. Thus it appears that for long tails one might prefer using the Huber scale version. For the double exponential the α-trimmed mean appears to trounce the Huber scale version for $k \geq 7$, $\alpha \leq .248$. However, when we tried using the mean absolute deviation as a scale estimate, the results were reversed. That is, the Huber scale version easily won for all k and α computed. This suggests that with judicious choice of scale, the Huber scale version might perform better in a variety of situations. It would be interesting to try other scale estimates or use a simultaneous estimation scheme such as Huber's Proposal 2 (see Huber (1964), p. 96).

TABLE 1.

$F = \Phi$, standard normal

k	α=F(-k)	$A_k(F)$	$A_S(F)$	α-trimmed mean	Huber	Huber-Scale version
0.1	0.460	1.492	1.360	16.871	17.242	16.306
0.2	0.421	1.423	1.360	7.303	7.555	6.864
0.3	0.382	1.362	1.360	4.256	4.415	3.932
0.4	0.345	1.309	1.360	2.816	2.902	2.597
0.5	0.309	1.263	1.360	2.006	2.035	1.877
0.6	0.274	1.222	1.360	1.500	1.484	1.450
0.7	0.242	1.187	1.360	1.164	1.113	1.179
0.8	0.212	1.156	1.360	0.928	0.850	0.997
0.9	0.184	1.130	1.360	0.757	0.658	0.868
1.0	0.159	1.107	1.360	0.630	0.515	0.772
1.1	0.136	1.088	1.360	0.532	0.405	0.696
1.2	0.115	1.072	1.360	0.456	0.320	0.632
1.3	0.097	1.058	1.360	0.396	0.254	0.576
1.4	0.081	1.047	1.360	0.347	0.202	0.525
1.5	0.067	1.037	1.360	0.307	0.160	0.477
1.6	0.055	1.029	1.360	0.274	0.127	0.431
1.7	0.045	1.023	1.360	0.247	0.100	0.386
1.8	0.036	1.018	1.360	0.223	0.079	0.344
1.9	0.029	1.014	1.360	0.203	0.062	0.303
2.0	0.023	1.010	1.360	0.186	0.048	0.264
2.1	0.018	1.008	1.360	0.171	0.037	0.228
2.2	0.014	1.006	1.360	0.158	0.029	0.194
2.3	0.011	1.004	1.360	0.146	0.022	0.164
2.4	0.008	1.003	1.360	0.136	0.017	0.136
2.5	0.006	1.002	1.360	0.127	0.013	0.112

TABLE 2.

$F = .9\,\Phi + .1\Phi(x/10)$, contaminated normal

k	α=F(-k)	$A_k(F)$	$A_S(F)$	α-trimmed mean	Huber	Huber-Scale version
0.1	0.464	1.811	1.464	22.736	23.174	22.003
0.2	0.428	1.737	1.464	10.018	10.301	9.441
0.3	0.393	1.672	1.464	5.959	6.120	5.524
0.4	0.359	1.618	1.464	4.038	4.099	3.729
0.5	0.326	1.571	1.464	2.957	2.936	2.756
0.6	0.294	1.532	1.464	2.284	2.195	2.176
0.7	0.265	1.501	1.464	1.840	1.692	1.806
0.8	0.237	1.476	1.464	1.534	1.335	1.560
0.9	0.212	1.457	1.464	1.319	1.073	1.388
1.0	0.189	1.443	1.464	1.167	0.876	1.264
1.1	0.168	1.435	1.464	1.060	0.724	1.172
1.2	0.149	1.432	1.464	0.989	0.607	1.102
1.3	0.132	1.433	1.464	0.949	0.514	1.048
1.4	0.117	1.439	1.464	0.936	0.440	1.006
1.5	0.104	1.448	1.464	0.950	0.381	0.975
1.6	0.093	1.462	1.464	0.994	0.334	0.953
1.7	0.083	1.479	1.464	1.074	0.296	0.940
1.8	0.075	1.499	1.464	1.201	0.265	0.935
1.9	0.068	1.523	1.464	1.389	0.241	0.939
2.0	0.063	1.550	1.464	1.663	0.222	0.951
2.1	0.058	1.579	1.464	2.061	0.206	0.972
2.2	0.054	1.611	1.464	2.640	0.194	1.001
2.3	0.051	1.646	1.464	3.487	0.185	1.037
2.4	0.048	1.683	1.464	4.735	0.178	1.081
2.5	0.046	1.723	1.464	6.583	0.173	1.130

TABLE 3.

$F = t_3$, t distribution with 3 degrees of freedom

k.	α=F(-k)	$A_k(F)$	$A_S(F)$	α-trimmed mean	Huber	Huber-Scale version
0.1	0.463	1.768	1.612	21.955	22.334	21.202
0.2	0.427	1.700	1.612	9.771	9.965	9.198
0.3	0.392	1.645	1.612	5.917	5.963	5.499
0.4	0.358	1.602	1.612	4.110	4.038	3.823
0.5	0.326	1.570	1.612	3.103	2.934	2.923
0.6	0.295	1.546	1.612	2.484	2.232	2.388
0.7	0.267	1.530	1.612	2.078	1.756	2.046
0.8	0.241	1.522	1.612	1.801	1.417	1.816
0.9	0.217	1.519	1.612	1.607	1.167	1.653
1.0	0.196	1.521	1.612	1.468	0.976	1.534
1.1	0.176	1.527	1.612	1.369	0.828	1.443
1.2	0.158	1.537	1.612	1.299	0.711	1.372
1.3	0.142	1.550	1.612	1.250	0.616	1.315
1.4	0.128	1.565	1.612	1.218	0.539	1.267
1.5	0.115	1.582	1.612	1.199	0.474	1.227
1.6	0.104	1.601	1.612	1.191	0.420	1.192
1.7	0.094	1.621	1.612	1.192	0.375	1.161
1.8	0.085	1.642	1.612	1.201	0.336	1.133
1.9	0.077	1.664	1.612	1.216	0.302	1.108
2.0	0.070	1.686	1.612	1.237	0.273	1.084
2.1	0.063	1.709	1.612	1.263	0.248	1.062
2.2	0.058	1.731	1.612	1.294	0.225	1.041
2.3	0.052	1.754	1.612	1.329	0.206	1.021
2.4	0.048	1.776	1.612	1.368	0.188	1.002
2.5	0.044	1.799	1.612	1.410	0.173	0.983

TABLE 4.

F = D-EX, double exponential with density $f(x) = \frac{1}{2}e^{-|x|}$

k	α=F(-k)	$A_k(F)$	$A_S(F)$	α-trimmed mean	Huber	Huber-Scale version
0.1	0.452	1.033	2.081	10.508	9.825	10.351
0.2	0.409	1.067	2.081	5.517	4.817	5.388
0.3	0.370	1.100	2.081	3.858	3.143	3.758
0.4	0.335	1.133	2.081	3.033	2.303	2.960
0.5	0.303	1.165	2.081	2.542	1.796	2.495
0.6	0.274	1.198	2.081	2.216	1.457	2.195
0.7	0.248	1.230	2.081	1.986	1.213	1.988
0.8	0.225	1.261	2.081	1.816	1.029	1.839
0.9	0.203	1.292	2.081	1.685	0.885	1.728
1.0	0.184	1.323	2.081	1.582	0.770	1.641
1.1	0.166	1.352	2.081	1.499	0.675	1.573
1.2	0.151	1.382	2.081	1.431	0.596	1.517
1.3	0.136	1.410	2.081	1.375	0.528	1.470
1.4	0.123	1.438	2.081	1.327	0.471	1.429
1.5	0.112	1.465	2.081	1.287	0.421	1.393
1.6	0.101	1.492	2.081	1.253	0.377	1.359
1.7	0.091	1.517	2.081	1.224	0.339	1.328
1.8	0.083	1.542	2.081	1.198	0.305	1.298
1.9	0.075	1.566	2.081	1.176	0.275	1.268
2.0	0.068	1.589	2.081	1.156	0.249	1.239
2.1	0.061	1.611	2.081	1.140	0.225	0.209
2.2	0.055	1.633	2.081	1.125	0.203	1.180
2.3	0.050	1.653	2.081	1.111	0.184	1.149
2.4	0.045	1.673	2.081	1.100	0.167	1.119
2.5	0.041	1.692	2.081	1.089	0.151	1.088

REFERENCES

[1] Andrews, D. F. et al. (1972). Robust Estimates of Location. Princeton Univ. Press.

[2] Boos, D. D. (1977a). The functional approach in statistical theory and robust inference. Unpublished dissertation, Florida State University.

[3] Boos, D. D. (1977b). A differential for L-statistics. FSU Statistics Report 416 (ONR Technical Report No. 116), Florida State University.

[4] Boos, D. D. and Serfling, R. J. (1977). On differentials, asymptotic normality and almost sure behavior of statistical functions, with application to M-statistics for location. FSU Statistics Report 415 (ONR Technical Report No. 115), Florida State University.

[5] De Wet, T. and Venter, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. Ann. Statist. 1 380-387.

[6] Cartan, H. (1971). Differential Calculus. Hermann, Paris.

[7] Collins, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. Ann. Statist. 4 68-85.

[8] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Statist. 3 1189-1217.

[9] Filippova, A. A. (1962). Mises theorem on the asymptotic behavior of functionals of empirical distribution functions and statistical applications. Theor. Probability Appl. 7 24-57.

[10] Ghosh, J. K. and Subramanyam, K. (1974). Second order efficiency of maximum likelihood estimators. Sankhyā Ser. A 36 325-358.

[11] Gregory, G. G. (1976). Large sample theory of differentiable statistical functions. Unpublished manuscript.

[12] Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. Ann. Statist. 5 110-123.

[13] Hampel, F. R. (1974). The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69 383-393.

[14] Huber, P. J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35 73-101.

[15] Jaeckel, L. A. (1971). Robust estimates of location: symmetry and asymmetric contamination. Ann. Math. Statist. 42 1020-1034.

REFERENCES (Continued)

[17] Rao, C. R. (1960). Apparent anomalies and irregularities of the method of maximum likelihood. Proceedings of the 32nd Session of the Int. Stat. Inst., Tokyo. Reprinted with discussion in Sankhyā 24 73-102.

[18] Rao, C. R. (1961). Asymptotic efficiency and limiting information. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1 531-546.

[19] Rao, C. R. (1962). Efficient estimates and optimum inference procedure in large samples, with discussion. J. Royal Statist. Soc. Ser. B 24 46-63.

[20] Rao, C. R. (1963). Criteria of estimation in large samples. Sankhyā Ser. A 25 189-206.

[21] von Mises, R. (1947). On the asymptotic distributions of differentiable statistical functions. Ann. Math. Statis. 18 309-348.

[22] Zabreyko, P. P. et al. (1975). Integral Equations - a Reference Text. Noordhoff International Publishing, Leyden.