

MINIMUM DISTANCE AND ROBUST ESTIMATION

by

William C. Parr
William R. Schucany

Technical Report No. 136
Department of Statistics ONR Contract

October 5, 1979

Research Sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280

Reproduction in whole or in part is permitted
for any purpose of the United States Government

This document has been approved for public release
and sale; its distribution is unlimited

DIVISION OF MATHEMATICAL SCIENCES
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Abstract

Robust and consistent estimation of the location parameter of an asymmetric distribution and general, non-location and scale parameter estimation problems have been vexing problems in the history of robustness studies. The minimum distance (MD) estimation method is shown to provide a heuristically reasonable mode of attack for these problems which also leads to excellent robustness properties. Both asymptotic and Monte Carlo results for the familiar case of estimation of the location parameter of a symmetric distribution support this proposition, showing MD-estimators to be competitive with some of the better estimators thus far proposed.

Key Words: Robust estimation; Minimum distance; Non-location and scale problems; Influence curve; Swindle.

William C. Parr is Assistant Professor, Institute of Statistics, Texas A&M University, College Station, Texas 77843. William R. Schucany is Professor, Department of Statistics, Southern Methodist University, Dallas, Texas 75275.

1. INTRODUCTION AND HISTORY

A major concern in much recent statistical literature has been robust estimation, i.e. efficient or nearly efficient (at a model) estimation procedures which also perform well under moderate deviations from that model. Huber (1964) has proposed a class of M-estimators as solutions to a formally stated minimax problem of this type. However, as stated in Huber (1972, Sec. 12.3) and Huber (1977), problems occur when the attempt is made to extend these methods (highly successful when invariance and symmetry properties are present) to shape or truncation parameter models. Thus, there is a need for procedures which extend easily to the more difficult situations.

Wolfowitz (1957) published a fundamental paper outlining the minimum distance method, proving a consistency result, and giving a number of intriguing examples of its use. Interestingly, the motivation for his work was the existence of complex estimation problems, then unsolved via other methods. Knüsel (1969) examined the relationship of robustness considerations to the method of minimum distance (henceforth called the MD-method). For the particular discrepancy function studied most closely (which apparently requires numerical integration for its evaluation) he showed that his D-estimators belong to the class of

M-estimators. Littell and Rao (1975) and Rao, Schuster, and Littell (1975) have considered in some detail the use of the Kolmogorov distance for MD-estimation, emphasizing the two-sample shift problem but also addressing and obtaining results for the one sample location case. Holm (in the discussion of Bickel (1976)) has suggested MD-estimation as being the most natural method for some robustness problems, and a recent paper by Easterling (1976) approaches MD-estimation from the point of view of consonance regions in order to incorporate goodness-of-fit considerations directly into the problem of parameter estimation.

2. NOTATION AND DEFINITIONS

Several measures of the discrepancy between an empirical distribution function and a theoretical one are of special interest in this work. In the following we let $G_n(\cdot)$ denote the empirical distribution function based upon a random sample of size n from the (possibly unknown) true distribution function $G(\cdot)$, and $\Gamma = \{F_\theta(\cdot), \theta \in \Omega\}$ where Ω is some parameter space. Most of the discrepancies considered here are in use as goodness-of-fit statistics based upon the empirical distribution function (for surveys see Stephens (1974), Sahler (1968)). Let K and L denote two distribution functions with support a (common)

subset of R . A list of some measures of interest follows:

$$i) D_{\psi}(K, L) = \sup_{x \in R} |K(x) - L(x)| \psi(L(x)), \text{ the weighted}$$

Kolmogorov distance, with the uniform weighting function

$\psi(\cdot) \equiv 1$ of special interest.

$$ii) W_{\psi}^2(K, L) = \int_{-\infty}^{\infty} (K(x) - L(x))^2 \psi(L(x)) dL(x), \text{ the weighted}$$

Cramér-von Mises distance with the special weight functions of interest

a) $\psi(\cdot) \equiv 1$ yielding the Cramér-von Mises statistic

$$W^2(K, L)$$

$$b) \psi(u) = \frac{1}{u(1-u)}, \quad 0 < u < 1 \text{ yielding the}$$

Anderson-Darling statistic $A^2(K, L)$, and

$$c) \psi(u) = 1, \quad \epsilon < u < 1 - \epsilon$$

$$= 0, \quad \text{otherwise}$$

for some ϵ with $0 < \epsilon < \frac{1}{2}$ yielding a trimmed

Cramér-von Mises distance as suggested by Anderson and Darling (1952).

$$iii) V(K, L) = \sup_{-\infty < a < b < \infty} |(K(b) - K(a)) - (L(b) - L(a))|,$$

Kuiper's maximal interval probability distance.

(2.1) is attained, and that there is no guarantee that the infimum will be attained in the interior of Ω . Thus, we are forced to make the following our general definition of a sequence of MD-estimators.

Definition 2.1. A sequence of random variables $\{T_n\}_{n=1}^{\infty}$ is a sequence of asymptotic minimum distance estimators based on $\{G_n\}_{n=1}^{\infty}$ with respect to $\delta(\cdot, \cdot)$ and Γ if

$$i) \quad T_n \in \Omega \text{ for all } n \geq 1$$

and

$$ii) \quad \text{There exists a nonnegative function } K(n) \text{ with}$$

$$\lim_{n \rightarrow \infty} K(n) = 0 \text{ such that}$$

$$\delta(G_n, F_{T_n}) \leq \inf_{\theta \in \Omega} \delta(G_n, F_{\theta}) + K(n) \text{ for all } n \geq 1.$$

Similar structure has been used in this setting by Wolfowitz (1957) and Sahler (1970). The following consistency theorem holds by a straightforward argument.

Theorem 2.1. With all notation as above, if $\{T_n\}_{n=1}^{\infty}$ is a sequence of asymptotic MD-estimators based on $\{G_n\}_{n=1}^{\infty}$ with respect to $\delta(\cdot, \cdot)$ and the model Γ , and δ , G , and Γ satisfy:

$$i) \quad \text{for any sequence } \{H_n\}_{n=1}^{\infty},$$

$$\sup_x |H_n(x) - G(x)| \rightarrow 0 \text{ implies } \delta(H_n, F_{\theta}) \rightarrow \delta(G, F_{\theta})$$

uniformly over Ω .

ii) there is a point $\theta_0 \in \Omega$ such that

$$\inf_{\theta \in \Omega} \delta(G, F_\theta) = \delta(G, F_{\theta_0}) ,$$

iii) $\lim_{k \rightarrow \infty} \delta(G, F_{\theta_k}) = \delta(G, F_{\theta_0})$ implies $\lim_{k \rightarrow \infty} \theta_k = \theta_0$,

then $\lim_{n \rightarrow \infty} T_n = \theta_0$ with probability 1.

Some points worthy of note are the following:

- 1) This result is simply a statement of sufficient conditions for continuity of the functional $\theta_n = T(G_n)$ (with the Kolmogorov metric on the space of distribution functions) at G , which has been considered as a robustness property in itself (Bickel and Lehmann (1975a), Fu (1976), Hampel (1971), with use of the Prokhorov norm).
- 2) Sequences $\{G_n\}_{n=1}^\infty$ of functions other than empirical distribution functions are covered by the proof of the theorem. This is useful for a differential-type approach to the demonstration of asymptotic normality.
- 3) The major condition is i), requiring uniformity of the convergence over Ω . The theorem is presented in this fashion as most conducive to intuitive insight. The condition can of course be easily relaxed to local uniformity of the convergence. The conditions incidentally cover most cases of location parameter estimation (scale known) and are easily verified (especially

in the correct projection family case). Condition ii) merely specifies the uniqueness of the value $\theta_0 \in \Omega$ minimizing $\delta(G, F_{\theta})$, while iii) requires that the parametrization of Γ (and choice of $\delta(\cdot, \cdot)$) be sensible - that in order to get $\delta(G, F_{\theta})$ arbitrarily close to $\delta(G, F_{\theta_0})$, one must take θ sufficiently close to θ_0 . A similar theorem was published in Wolfowitz (1957) for a particular choice of $\delta(\cdot, \cdot)$.

MD-estimators share an invariance property with maximum-likelihood estimators in that $\hat{g}(\theta) = g(\hat{\theta})$, e.g. that an MD-estimator of μ^2 for a $N(\mu, \sigma^2)$ population is thus $(\hat{\mu})^2$, where $\hat{\mu}$ denotes an MD-estimator for μ . Thus, MD-estimation is invariant to choice of the function $g(\theta)$ of the point $\theta \in \Omega$ to be estimated, contrary to the case for UMVU estimation methods. It operates in a manner analogous to maximum likelihood methods in simply selecting a "best approximating distribution" from those in the model. (See Fisher (1973, p.146) in regard to the desirableness of this property.)

3. LOCATION PARAMETER ESTIMATION

3.1 Symmetric Parent, scale known.

In this section we let Γ be a translation family of symmetric continuous distribution, i.e. $\Gamma = \{F_{\theta} : F_{\theta}(x) = F(x - \theta), -\infty < \theta < \infty, -\infty < x < \infty, \text{ and } F(x) = 1 - F(-x), -\infty < x < \infty\}$ and assume also that G , the sampled distribution, is symmetric.

Also let G_n denote the empirical distribution function for a random sample of size n from G .

Although influence curves as in Hampel (1974) can be easily derived in the general case, we give explicit solutions only for the case $G \in \Gamma$, taking $G = F_0 = F$ without loss of generality. In fact, the case $G \notin \Gamma$ seems to possess little if any meaning or significance unless scale is estimated simultaneously. Influence curves for minimum D and V estimators are not obtainable by our methods (and may well not exist). The MD-estimators of location obtained by using D and V as discrepancies are not even asymptotically normal in the simplest cases. (Littell and Rao (1975), Böhthausen (1977) show asymptotic equivalence of the MD-estimator based on D to a complicated (and clearly nonnormal) function of a Brownian bridge.) We consider the use of discrepancies of the form $Z_{a,b}^2$ as a rather general class including the Cramér-von Mises, Watson's U^2 , and Chapman discrepancies as special cases. The usual implicit differentiation yields as the influence curve for the derived estimator

$$IC_{T,F}(c) = \left\{ a \int_{-\infty}^{\infty} (F(x) - \delta_c(x)) f^2(x) dx + b \left[\int_{-\infty}^{\infty} (F(x) - \delta_c(x)) f(x) dx \right] \int_{-\infty}^{\infty} f^2(x) dx \right\} \left[a \int_{-\infty}^{\infty} f^3(x) dx + b \left(\int_{-\infty}^{\infty} f^2(x) dx \right)^2 \right]^{-1},$$

$$-\infty < c < \infty,$$

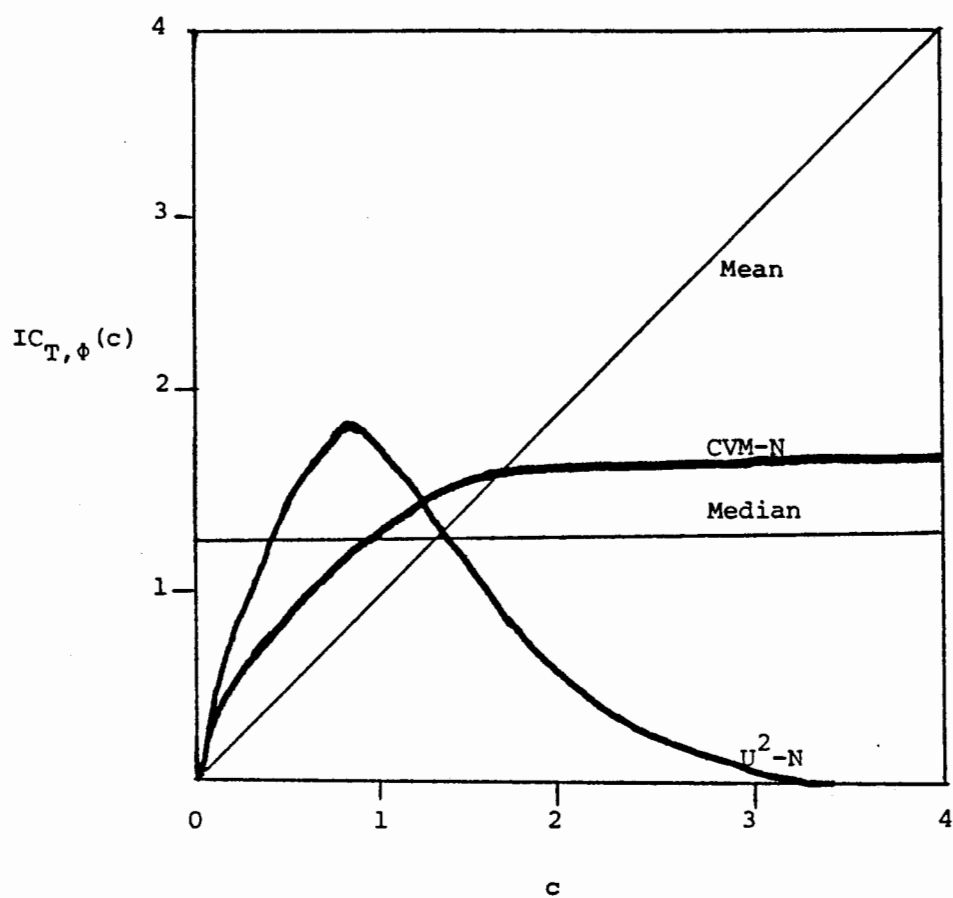
which is a valid expression for all $a \geq 0$, $a + b \geq 0$ with one inequality strict. At the normal parent, $G = F = \Phi$,

$$IC_{T,\Phi}(c) = \begin{cases} \sqrt{3\pi} (\Phi(\sqrt{2c}) - \frac{1}{2}), & -\infty < c < \infty & \text{for } a = 1, b = 0 \\ & \text{(wrt the Cramér-von Mises discrepancy)} \\ \frac{\sqrt{3\pi} (\Phi(\sqrt{2c}) - \Phi(c))}{1 - \frac{\sqrt{3}}{2}}, & -\infty < c < \infty & \text{for } a = 1, b = -1 \\ & \text{(wrt the Watson's } U^2 \text{ discrepancy)}. \end{cases}$$

Note in Figure 3.1 that the minimum U^2 is redescending at the normal and other symmetric models as long as $G \in \Gamma$. It should be mentioned that MD-estimation of location parameters using U^2 as a discrepancy measure is not being advocated here, but simply being used as an analytically simple and illustrative example. The fact that U^2 is more powerful (as a goodness-of-fit test) against alternatives involving a scale shift than against location shifts (see Stephens (1974)) serves as an indicator that it should be a poor choice as a discrepancy measure for location estimation, but a good one for scale estimation.

(Figure 3.1 about here)

Table 3.1 contains gross-error-sensitivities and asymptotic variances at the normal parent for these two estimators and some others as tabled in Hampel (1974). The low gross-error-

FIGURE 3.1 Graph of Influence Curves at Normality^a

^a The CVM-N and U^2-N are as given in Section 3.1. Only the positive half of the influence curves is displayed, as all four are odd functions.

sensitivity of the minimum W^2 estimator (second only to the median among those tabled by Hampel) is noteworthy, as is the (expected) high variance of the minimum U^2 estimator. It is somewhat curious that projection onto the normal parent via a goodness-of-fit distance should lead to estimators with any robustness at all. The basic principle seems to be that robustness is due to measuring the discrepancy between observed data and model in "probability-type" units. In cases such as the Anderson-Darling discrepancy, where the weight given to deviations in probability units from the model is high in the tails of the distribution, drastic sensitivity to incorrect tail-width specification can be expected. Typical measures of interest, as exemplified by those listed in Section 2 (excluding unboundedly weighted Kolmogorov or Cramér-von Mises discrepancies) assign either equal or less weight to discrepancies between the model and the data in regions of low probability content for the model. In fact, the Cramér-von Mises discrepancy drastically downweights discrepancies in the tails. The "trimmed" versions of the weighted Cramér-von Mises discrepancy are in fact designed to further minimize the effect of extreme observations. The V discrepancy, designed for the goodness-of-fit problem on the circle, weights all discrepancies equally.

(Table 3.1 about here)

TABLE 3.1

Asymptotic Variances and Gross-Error-Sensitivities

<u>Estimator</u>	<u>σ^2</u>	<u>γ^*</u>
CVM - N	1.095	1.53
U^2 - N	1.869	1.90
M	1.000	∞
25A	1.026	1.86
H(1.5)	1.037	1.73
50%	1.571	1.25
10% Trim	1.060	1.60
H/L	1.047	1.77

Entries in table:

σ^2 = Asymptotic variance

γ^* = Gross-error-sensitivity

Estimators are a minimum Cramér-von Mises estimator (CVM - N) and a minimum Watson's U^2 estimator (U^2 - N), both projecting onto the normal location family, the mean (M), median (50%), and several estimators are tabled in Hampel (1974). Note that all values in the table are at the $N(0,1)$ parent.

Asymptotic normality of these estimators can be established by using techniques similar to that used by Boos and Serfling (1977) for M-estimators. Briefly put, it is necessary only to show

$$|T[G_n] - T[G] - H(G_n)D_T[G_n - G]| = o(\|G_n - G\|_\infty), \quad (3.1)$$

where $\{G_n\}_{n=1}^\infty$ is a sequence such that

$$\|G_n - G\|_\infty = \sup_x |G_n(x) - G(x)| \rightarrow 0,$$

$T[\cdot]$ represents the estimator under consideration as a functional on (an appropriate subset of) the space of univariate distribution functions, and $H(\cdot)$ a functional on the same space such that

$$\lim_{\|G_n - G\|_\infty \rightarrow 0} H(G_n) = 1, \text{ and } D_T(G_n - G) \text{ is linear, i.e. there}$$

exists a function $\psi(\cdot)$ such that

$$D_T[G_n - G] = \int_{-\infty}^{\infty} \psi(x) d(G_n - G)(x)$$

for the set of $G_n - G$ corresponding to the above collections of G_n . The approach of Boos and Serfling can be closely paralleled for the most part, leading to the following theorem. Note that in the above, $\psi(x) = IC_{T,G}(x) + \text{an arbitrary constant}$.

Theorem 3.1 Let Γ , G_n , G , and T be defined as above, with F and G symmetric. If T is a MD-estimator with respect to the discrepancy $Z_{a,b}^2$ and Γ and

$$i) \quad T[G_n] \xrightarrow{P} \theta_0$$

and

$$ii) \quad \int_{-\infty}^{\infty} f^3(x) dx < \infty, \quad \int_{-\infty}^{\infty} |f'(x)| dx < \infty$$

then

$$\lim_{n \rightarrow \infty} P[\sqrt{n}(T[G_n] - T[G]) \leq z] = \Phi\left(\frac{z}{\sigma_T}\right),$$

where $\sigma_T^2 = \int_{-\infty}^{\infty} IC_{T,G}^2(x) dG(x) > 0$. A proof is sketched in the Appendix.

Some points for comment are the following:

- 1) Apart from the conditions for consistency (Theorem 2.1) the burden of the regularity conditions is carried by the projection family Γ , over which the statistician has control, rather than the unknown distribution function G .
- 2) Equation (3.1) is equivalent in most cases to

$$T[G_n] = T[G] + \int_{-\infty}^{\infty} IC_{T,G}(x) d(G_n - G)(x) + o(\|G_n - G\|_{\infty}),$$

giving the asymptotic equivalence which justifies (for the asymptotic case) the usual heuristic interpretations of the influence curve. This resultant asymptotic expansion thus extends the normality results of Sahler (1970).

3) Since differentiability implies continuity and (3.1) is equivalent to

$$\lim_{n \rightarrow \infty} \frac{|T[G_n] - T[G] - \int_{-\infty}^{\infty} \psi(x) d(G_n - G)(x)|}{\|G_n - G\|_{\infty}} = 0$$

for most $\psi(\cdot)$ (all considered here), a Frechet-type differentiability result for estimators derived from $Z_{a,b}^2$ is given which could be used for a new definition of robustness, somewhat parallel to but more stringent than that of Beran (1977a), to be considered in a future paper.

4) Identical results hold when scale is also unknown and $G \in \Gamma$, with the case of unknown scale and $G \notin \Gamma$ not yet fully resolved by the authors.

5) Parr and DeWet (1979) show asymptotic normality of $T[G_n]$ in the correct model case with weighted Cramér-von Mises discrepancies for general parameters. The proof would easily extend to a weighted version of $Z_{a,b}^2$.

6) The symmetry of F and G was specified only to simplify the notation. For $G \in \Gamma$ this restriction may be omitted. For $G \notin \Gamma$ and in the absence of symmetry it typically will suffice for $\delta(G, F_{\theta})$ to have a unique minimum θ_0 at which it is suitably differentiable.

3.2 Scale Unknown Cases.

Typically, all the remarks in Section 3.1 hold for the scale unknown, location estimation problem when G and F are both symmetric. Then, the parameter θ is two-dimensional and location and scale are estimated simultaneously. Here, the scale estimator $S(G_n)$ is consistent and asymptotically independent of the location estimator (see Huber (1972) for a related remark regarding M -estimators) and thus the asymptotic properties of the location estimator are the same as if scale were known, i.e. $S(G_n) = S(G)$ for all G_n .

4. MONTE CARLO RESULTS FOR LOCATION ESTIMATION

A Monte Carlo investigation of the performance characteristics of MD-estimators over a wide variety of symmetric distributions in the location estimation problem (scale unknown) is reported in this section. This case is the best studied and understood, permitting direct comparisons with Monte Carlo studies of other proposals. All computations were performed on the CDC Cyber 72 at Southern Methodist University.

It was felt that such a study was in order for several reasons - i) to relate the large-sample theory for MD-estimators to the practical small-sample situations likely to be encountered, ii) to explore the behavior of the MD-estimators based on

sup-type discrepancies, for which the large-sample theory is incomplete, and iii) to bolster the authors' argument that MD-estimators are easily applicable and may well be good for more complex parameter estimation problems.

The distributions G , for which results are reported (with abbreviated notation in parentheses) include the standard normal ($N(0,1)$), t -distributions with 8, 4, 2, and 1 degrees of freedom ($T(8)$, $T(4)$, $T(2)$, $T(1)$), the Laplace distribution (LAP), fixed proportion (3:1) mixtures of standard normals and slash (quotients of standard normals and independent uniforms) (3N1S), fixed proportion mixtures (9:1) of standard normals and normals with standard deviations of 3 and 10 (10% 3N and 10% 10N respectively) and a fixed (equal) proportion mixture of standard normals and uniform variates with mean 0 and variance 1 (50%U*). All distributions except the last have tailweight greater than or equal to that of the normal. Generation of the normal variates was done via the polar method, with all required uniforms generated by a multiplicative congruential method. Chi-square variates were generated via the IMSL routine GGCSS. Primary attention is focused upon sample size $n = 20$, with a subset of the above configurations examined for $n = 10$. The Princeton Swindle (Gross (1973)) was employed to reduce Monte Carlo variability for all but the distribution 50%U*.

This highly efficient swindle is based upon variates of the form $X = Z/Y$ where $Z \sim N(0,1)$ and Y are independent. Unfortunately the kurtosis, $K(X)$, of such variates satisfies $K(X) \geq K(Z)$, regardless of the distributions of Z and Y (subject to the existence of the relevant moments). Also, the swindle does not appear to extend easily to numerators other than the normal. Thus, the ideas of this method seem to be presently unusable for short-tailed populations in general. (See Parr (1979) for an extension to the uniform case.) All results quoted are based upon 1000 repetitions.

Table 4.1 is a glossary for the estimators for which performance measures are given later. MD-estimators with "fixed" scale estimation utilize the (properly scaled for the projection model) sample interquartile range as a scale estimate and minimize the discrepancy over choice of the location parameter via the IMSL Fibonacci-type minimization routine ZXFIB. The estimate is taken to be the sample median when the minimizing value falls outside of the first and third quartiles. This routine does not require specification of the derivatives of the objective function with respect to the parameters, and thus is not the most efficient choice in general. The authors chose it, however, to demonstrate the reasonable practicality of the MD - method - not requiring special routines beyond a

function to compute $F_{\theta}(\cdot)$, one to compute $\delta(F_{\theta}, G_n)$, and an omnibus minimization routine. In spite of this, the routine converged to within an accuracy of .005 for $T[G_n]$ rather quickly for CVM-N, converging in an average of .115 seconds (typically 12-14 iterations) for the $N(0,1)$ parent, .120 seconds (12-14 iterations) for a Cauchy parent, and .115 seconds (12-14 iterations) for a Laplace parent. These compared to typical times for the M-estimators of .005-.006 for all three parents. The average cost of any single estimator studied was less than one cent at the current rates for the SMU Cyber 72. Subsequent experimentation with rational function approximations to the normal cumulative distribution function reduced the average times for the MD-estimators by a factor of two from those times quoted above.

MD-estimators with "simultaneous" scale estimation initialize the location and scale parameters at the median and rescaled interquartile range, minimizing the discrepancy jointly in the two parameters via the IMSL quasi-Newton ZXMIN algorithm. This routine approximates the derivatives of $\delta(G_n, F_{\theta})$ with respect to θ numerically. No initial estimate of the Hessian matrix is required. As before ZXMIN is a good omnibus minimization routine chosen to demonstrate the ease of implementation of MD-estimators. For this routine, CVM-N converged in average times roughly twice those for fixed scale estimation.

For the Cramér-von Mises type discrepancies used herein, (including the trimmed ones) verification that Theorem 3.1 holds is a matter of showing that the 1) model density obeys (ii) of that theorem (trivial for the models considered in this Monte Carlo study), 2) $\sigma_T^2 > 0$, and 3) the consistency condition. The other MD-estimators, based upon the Kolmogorov and Kuiper discrepancies, do not have asymptotically normal distributions, as mentioned in Section 3.1.

(Table 4.1 about here)

The outer mean OM (the average of the 25% largest and 25% smallest values in the sample) was included to demonstrate the drastic inefficiency of existing proposed robust estimators for short-tailed situations. All other estimators have mnemonics as in Andrews, et.al. (1972) (H10, H15, H20, 12A, 17A, 21A, 22A, 25A, HGP, GAS, 50%, M) and are computed as in routines contained therein. The Hampels and Hubers were included as families including some of the best and most studied estimators in the literature.

Entries in Tables 4.2a and b are 20 times (estimated variance of estimator). An approximate standard error for each entry in a given column can be obtained as $S_{\sigma_T^2} = .0447$ (Entry-a), where a is the value in the last row of that column and is related to the savings due to the swindle. More digits than are often significant are included since the blocking effect due

TABLE 4.1

Mnemonic	Scale Estimate	Discrepancy	Projection Model
CVM-N	Fixed	Cramér-von Mises	Normal
TCVM-N-.10	"	Trimmed Cramér-von Mises ($\epsilon=.10$)	"
TCVM-N-.20	"	" ($\epsilon=.20$)	"
CVM-T(4)	"	Cramér-von Mises	t with 4 degrees of freedom
KS-N	"	Kolmogorov-Smirnov	Normal
V-N	"	Kuiper's V	"
U ² -N	"	Watson's U ²	"
SCVM-N	Simultaneous	Cramér-von Mises	Normal
STCVM-N-.10	"	Trimmed Cramér-von Mises ($\epsilon=.10$)	"
STCVM-N-.20	"	" ($\epsilon=.20$)	"
SKS-N	"	Kolmogorov-Smirnov	"
SV-N	"	Kuiper's V	"
SU ² -N	"	Watson's U ²	"

to using the same samples across all estimators makes qualitative comparisons of different estimators at the same distribution more precise. It should be mentioned that the swindle, which is responsible for $a > 0$ in all but the short-tailed case, produces more precise variance estimates for more efficient estimators and for more near-normal distributions. Table 4.3 contains similar results at $n = 10$ for a smaller set of distributions. Comparisons with both exact theoretical values and previous Monte Carlo work bolster faith in the estimated variances and their approximate standard errors.

(Tables 4.2a, b and 4.3 about here)

Several points are worthy of note based upon a general inspection of the tables. Distributions not examined in the Princeton Robustness Study (PRS) are 50%U*, T(8), T(4), and T(2). In general, the MD-estimators seem to fare extremely well for all but the most drastic heavy-tailed alternatives to normality, in comparison with even the best of the M-estimators considered here. A perusal of the relative behavior of MD-estimators using fixed or simultaneous scale estimation reveals the simultaneous estimation of scale to be profitable when the sampled distribution G is not near the projection family Γ , but in fact a liability otherwise.

4.2a MONTE CARLO VARIANCES FOR LOCATION ESTIMATORS

n = 20

Population Estimator					
	N(0,1)	T(8)	T(4)	T(2)	T(1)
Fixed Scale					
CVM-N	1.0595	1.2383	1.4621	2.1044	4.3926
TCVM-N-.10	1.0648	1.2415	1.4605	2.0891	4.3156
TCVM-N-.20	1.0913	1.2550	1.4526	2.0344	4.0872
CVM-T(4)	1.0912	1.2531	1.4497	1.9956	3.8347
KS-N	1.0871	1.3125	1.5835	2.4439	5.6664
V-N	1.9486	2.1242	2.2025	2.3516	3.1590
U ² -N	1.4052	1.5306	1.6297	1.8919	2.7428
Simultaneous Scale					
SCVM-N	1.0852	1.2478	1.4484	1.9950	3.6514
STCVM-N-.10	1.1107	1.2671	1.4519	1.9436	3.3410
STCVM-N-.20	1.1722	1.3122	1.4719	1.8943	3.0350
SKS-N	1.1257	1.2993	1.4996	2.1239	4.5855
SV-N	1.9960	2.1095	2.0761	2.1872	2.8649
SU ² -N	1.8270	1.8569	1.8462	1.9999	2.5863
M	1.0000	1.3126	2.0596	10.8307	*****
50%	1.4571	1.5897	1.7297	1.9861	2.7777
GAS	1.2102	1.3446	1.4939	1.8905	3.1305
OM	1.1754	1.8529	3.9725	35.7840	*****
HGP	1.0290	1.3180	1.6561	2.4014	3.7346
H10	1.0979	1.2571	1.4498	1.9902	3.7026
H15	1.0363	1.2390	1.5120	2.3343	5.7788
H20	1.0135	1.2520	1.5985	2.6966	8.5473
12A	1.2006	1.3275	1.4829	1.8908	2.7843
17A	1.1073	1.2764	1.4681	1.9584	3.0951
21A	1.0672	1.2599	1.4935	2.0814	3.4441
22A	1.0905	1.2899	1.5231	2.0949	3.4714
25A	1.0389	1.2499	1.5215	2.2007	3.8362
a	1.0000	1.0127	1.0256	1.0526	1.1111

4.2b MONTE CARLO VARIANCES FOR LOCATION ESTIMATORS

n = 20

Population Estimator	LAP	10%3N	10%10N	3NLS	50%U*
Fixed Scale					
CVM-N	1.4112	1.3091	1.4571	1.6174	1.2215
TCVM-N-.10	1.4023	1.3108	1.4522	1.6107	1.2275
TCVM-N-.20	1.3703	1.3177	1.4355	1.5914	1.3061
CVM-T(4)	1.3509	1.3244	1.4446	1.5934	1.3113
KS-N	1.5337	1.3832	1.5997	1.8484	1.0972
V-N	1.8125	2.1943	2.0868	2.2660	2.5710
U ² -N	1.4035	1.5530	1.4658	1.6537	1.8497
Simultaneous Scale					
SCVM-N	1.3599	1.3190	1.4400	1.5867	1.3020
STCVM-N-.10	1.3285	1.3384	1.4482	1.5893	1.3688
STCVM-N-.20	1.3039	1.3883	1.4860	1.6111	1.5216
SKS-N	1.3906	1.3731	1.5074	1.7037	2.8267
SV-N	1.6825	2.1443	1.9482	2.1866	1.3217
SU ² -N	1.5274	1.8497	1.6795	1.9054	2.9790
M	2.0450	1.7594	10.2602	*****	1.0158
50%	1.3553	1.6574	1.7422	1.9203	1.9867
GAS	1.3405	1.4204	1.5107	1.6482	1.6096
OM	3.8071	3.2054	33.9266	*****	.8674
HGP	1.6424	1.4532	1.7297	1.8731	1.0139
H10	1.3622	1.3203	1.4292	1.5789	1.3467
H15	1.5386	1.2973	1.4763	1.6932	1.1277
H20	1.6720	1.3407	1.6501	1.9692	1.0506
12A	1.3415	1.3752	1.3481	1.5436	1.5581
17A	1.3857	1.3138	1.2816	1.4788	1.3165
21A	1.4593	1.2929	1.2637	1.4820	1.1922
22A	1.5087	1.3062	1.2610	1.4968	1.2055
25A	1.5208	1.2952	1.2811	1.5165	1.1177
a	.5220	1.0975	1.1105	1.2018	0.0000

4.3 MONTE CARLO VARIANCES FOR LOCATION ESTIMATORS

n = 10

Population Estimator	N(0,1)	T(4)	Laplace	10%10N	T(2)	4N1S
Fixed Scale						
CVM-N	1.0783	1.4762	1.4588	1.4425	2.1444	1.4986
TCVM-N-.10	1.0838	1.4780	1.4547	1.4423	2.1325	1.4982
TCVM-N-.20	1.1089	1.4771	1.4293	1.4283	2.0894	1.4944
CVM-T(4)	1.1109	1.4750	1.4087	1.4338	2.0576	1.4944
KS-N	1.0796	1.5770	1.5386	1.5263	2.3515	1.6062
χ^2 -N	1.9145	2.1881	1.7790	2.0044	2.4558	2.1384
U^2 -N	1.4783	1.7256	1.5008	1.4974	2.0316	1.6529
Simultaneous Scale						
SCVM-N	1.0871	1.4678	1.4377	1.4332	2.1079	1.4877
STCVM-N-.10	1.1208	1.4770	1.4115	1.4431	2.0715	1.4984
STCVM-N-.20	1.1998	1.5206	1.3791	1.4835	2.0102	1.5560
SKS-N	1.1191	1.5278	1.4348	1.4740	2.1895	1.5516
SV-N	1.9993	2.1807	1.7893	1.9363	2.3932	2.1081
SU^2 -N	1.8501	2.0062	1.6419	1.6881	2.1940	1.8837
M						
50%	1.0000	2.0491	1.9407	11.1790	9.3433	2283.0178
GAS	1.4031	1.7117	1.4045	1.6630	2.0914	1.7393
OM	1.2288	1.5315	1.3755	1.4964	1.9823	1.5621
HGP	1.1081	3.2184	3.0410	27.0583	21.4474	6335.8547
H10						
H15	1.0411	1.7605	1.8899	1.6689	3.0374	1.9998
H20	1.1014	1.4736	1.4446	1.4174	2.1101	1.4817
12A	1.0339	1.5271	1.6122	1.4731	2.4550	1.5762
17A	1.0120	1.6230	1.7264	1.6600	3.0630	2.0553
21A	1.2264	1.5509	1.4130	1.3551	1.9482	1.4911
22A	1.1500	1.5009	1.4535	1.2868	2.0192	1.4250
25A	1.1044	1.5087	1.4952	1.2661	2.1189	1.4299
a	1.1496	1.5483	1.5410	1.2751	2.1237	1.4463
	1.0670	1.5174	1.5448	1.2779	2.2402	1.4596
	1.0000	1.0526	.6562	1.1099	1.1111	1.1565

Table 4.4 gives a smaller set of values more amenable to graphical presentation. The entries are, for $n = 20$, $E_i[j] = \hat{\text{Var}}(T_i \text{ at distribution } j) / \hat{\text{Var}}(\text{best } T \text{ at distribution } j)$, i.e., estimated (efficiency)⁻¹ relative to the (empirically determined) best estimator for that distribution. This adjusts for scale differences in the sampled populations, which for example avoids the difficult matter of rescaling $T(2)$ to be in some sense comparable with $N(0,1)$, thus permitting meaningful comparisons across distributions. Furthermore, based upon a first-order approximation, these entries should have smaller coefficients of variation than those in 4.2a and b since the numerators and denominators are highly correlated, both being estimates of the variances of fairly efficient estimators based upon the same data.

The generally good behavior of MD-estimators based upon CVM-type discrepancies when the sampled distribution is a t with moderate degrees of freedom stands out as before. Figure 4.1 plots $E_i[T(4)]$ versus $E_i[N(0,1)]$ for a number of the estimators considered. With this plotting system, good estimators will lie towards the bottom left of the plot. MD-estimators utilizing the Kolmogorov discrepancy are clearly inferior for moderately longtailed deviations from normality, while those using CVM-type discrepancies perform quite well both here and under normality.

Figure 4.2 gives a similar plot for a $E_1[T(2)]$ versus $E_1[N(0,1)]$ comparison. Here, M and MD-estimation methods seem to be equally good. For heavytailed symmetric distributions beyond the T(2) (Cauchy, Slash, or mixtures involving high proportions of Cauchy or Slash) the Hampels (particularly 12A) emerge as by far the best choice. The MD-estimators using V and U^2 being to exhibit some merit in these situations, in contrast to their disappointing behavior at the normal parent.

(Table 4.4, Figures 4.1, 4.2 about here)

In summary, the MD-estimators are quite competitive (while still not finely tuned) for all but the most drastic alternatives to normality. Furthermore, additional study may well reveal (as suggested by the behavior of CVM-T(4)) that the use of moderately or perhaps drastically heavytailed projection families Γ produces MD-estimators which work well for this case also. The behavior of KS(fixed scale) for 50% U^* suggests some hope for the shorttailed situation as well.

5. SUMMARY AND CONCLUSIONS

Both theoretical and Monte Carlo results have been given to suggest that MD-estimation is competitive with the better of the extant methods for the simple, symmetric-location estimation

4.4 EMPIRICAL DEFICIENCIES OF SELECTED ESTIMATORS

	N(0,1)	T(4)	10%3N	3N1S	T(2)
CVM-N	1.06	1.01	1.01	1.09	1.11
TCVM-N-.10	1.06	1.01	1.01	1.09	1.10
TCVM-N-.20	1.09	1.00	1.02	1.07	1.08
CVM-T(4)	1.09	1.00	1.02	1.08	1.06
KS-N	1.09	1.09	1.07	1.25	1.29
V-N	1.95	1.52	1.69	1.53	1.24
U ² -N	1.41	1.13	1.20	1.12	1.00
SCVM-N	1.09	1.00	1.02	1.07	1.06
STCVM-N-.10	1.11	1.00	1.03	1.07	1.03
STCVM-N-.20	1.17	1.02	1.07	1.09	1.00
SKS-N	1.13	1.04	1.06	1.15	1.12
SV-N	2.00	1.43	1.66	1.48	1.16
SU ² -B	1.83	1.27	1.43	1.29	1.06
H10	1.10	1.00	1.02	1.07	1.05
H15	1.04	1.04	1.00	1.14	1.23
H20	1.01	1.10	1.04	1.33	1.43
A12	1.20	1.02	1.06	1.04	1.00
A17	1.11	1.01	1.01	1.00	1.04
A21	1.07	1.03	1.00	1.00	1.10
A22	1.09	1.16	1.00	1.01	1.11
A25	1.04	1.05	1.00	1.03	1.16
M	1.00	1.42	1.36	*****	5.73
50%	1.46	1.19	1.28	1.30	1.05
Best	M	SCVM-N	21A	17A	12A

FIGURE 4.1 Plot of $E_i[T(4)]$ Versus $E_i[N(0,1)]$
for Selected Estimators

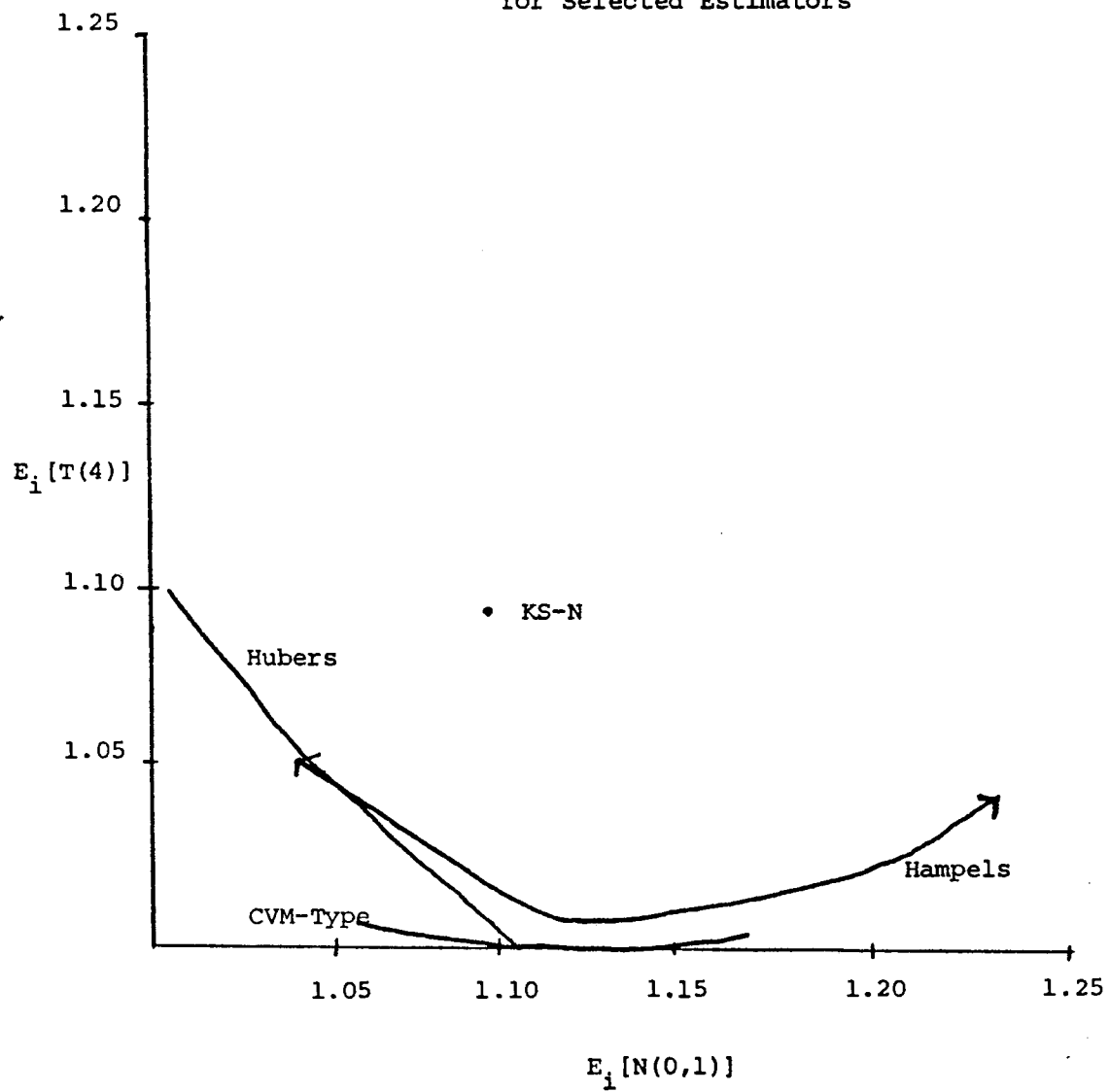
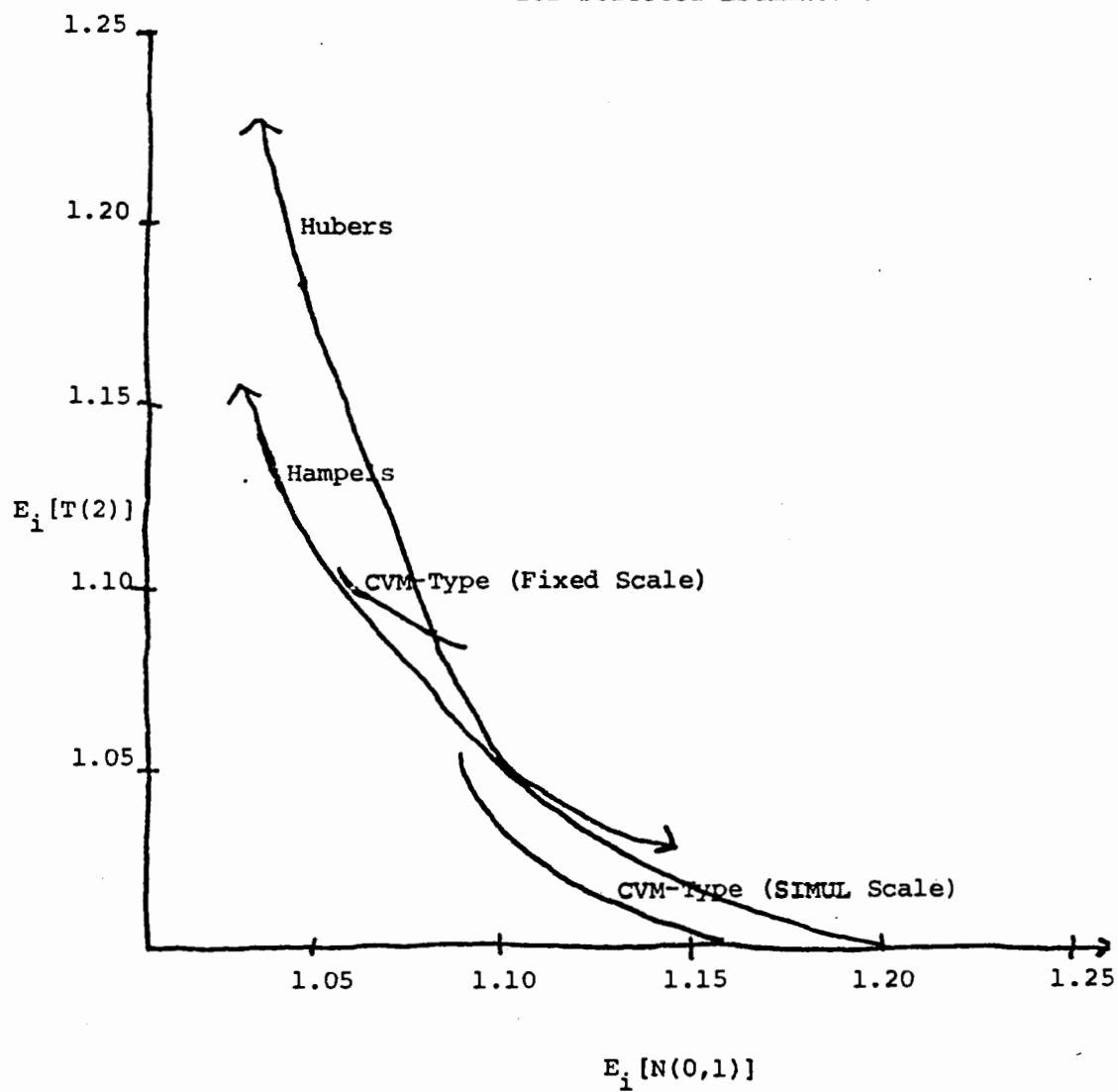


FIGURE 4.2 Plot of $E_i[T(2)]$ Versus $E_i[N(0,1)]$
for Selected Estimators



problem. MD-estimators seem to be, however, far easier to apply for more general estimation problems (strongly consistent and robust estimators are easy to derive and compute) than M-estimators, which can become quite intractable in situations in which these symmetry and invariance properties do not hold.

APPENDIX A

Herein we prove Theorem 3.1 for the case $G = F \in \Gamma$. The other case $G \notin \Gamma$ may be treated by a similar argument, but is not proven here due to space considerations.

Proof of Theorem 3.1

Define

$$\lambda_K(c) = \frac{\partial \delta(K, F_\theta)}{\partial \theta} \Big|_{\theta=c}, \quad -\infty < c < \infty$$

for all distribution functions K .

$T[G_n]$ is thus a root of $\lambda_{G_n}(T(G_n)) = 0$, where we assume a method of selecting a consistent root (for instance the one closest to the median). Define also the function (continuous at $t = T[F]$)

$$h(t) = \begin{cases} \frac{\lambda_F(t)}{t - T[F]}, & t \neq T[F] \\ \lambda'_F(T[F]), & t = T[F]. \end{cases}$$

Simple calculation for $\delta = Z_{a,b}^2$ yields

$$\lambda'_F(T[F]) = 2a \int f^3 d\mu + 2b(\int f^2 d\mu)^2.$$

Hence

$$T[G_n] - T[F] = \frac{\lambda_F(T(G_n))}{h(T(G_n))},$$

and $h(T(G_n)) \rightarrow \lambda'_F(T[F])$ with probability 1.

We desire to show that the following is legitimate as an expression for the differential,

$$\begin{aligned} D_T(G_n - F) &= \int IC_{T,F}(c) d(G_n - F) \\ &= \frac{2}{\lambda'_F(T[F])} \left[a \int (F - G_n) f^2 d\mu + b \left(\int f^2 d\mu \right) \left(\int (F - G_n) f d\mu \right) \right]. \end{aligned}$$

Define also $H(G_n) = \frac{\lambda'_F(T(F))}{h(T(G_n))}$, which converges to 1 with probability 1.

We may thus write

$$\begin{aligned} & |T[G_n] - T[F] - H(G_n)D_T(G_n - F)| \\ &= \left| \frac{1}{h(T(G_n))} \right| \left| \lambda'_F(T(G_n)) - 2(a \int (F-G)f^2 d\mu + b(\int f^2 d\mu)(\int (G-F)f d\mu)) \right| \end{aligned}$$

and must now demonstrate that this tends to zero faster than

$\|G_n - F\|_\infty$. The right hand side reduces after considerable mathematical manipulation to

$$\begin{aligned} & \left| \frac{1}{h(T(G_n))} \right| \left| 2a \int (F-G_n)(f_{T[G_n]}^2 - f^2) d\mu + \lambda'_{G_n}(T(G_n)) \right. \\ & + a \int [(G_n - F_{T[G_n]})^2 - (F - F_{T[G_n]})^2] f'_{T[G_n]} d\mu \\ & + 2b[(\int f^2 d\mu)(\int (F_{T[G_n]} - G_n) + (F - F_{T[G_n]})(f_{T[G_n]} - f) d\mu \\ & + 2b[(\int (G_n - F_{T[G_n]}) f_{T[G_n]} d\mu) \int (G_n - F_{T[G_n]}) f'_{T[G_n]} d\mu \\ & - (\int (F - F_{T[G_n]}) f_{T[G_n]} d\mu) \int (F - F_{T[G_n]}) f'_{T[G_n]} d\mu] \left. \right| \\ & \leq \left| \frac{1}{h(T(G_n))} \right| o(\|G_n - F\|_\infty) = o(\|G_n - F\|_\infty) \end{aligned}$$

by the triangle inequality and finiteness of $\int |f'| d\mu$. Thus,

$|T[G_n] - T[F] - H(G_n)D_T(G_n - F)| = o(\|G_n - F\|_\infty)$ and the theorem follows immediately by the Lindeberg-Levy version of the central limit theorem.

The case $G \notin \Gamma$ follows similarly, but with many more terms to be bounded in comparable fashion.

REFERENCES

- Anderson, T.W., and Darling, D. A. (1952), "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," Annals of Mathematical Statistics, 23, 193-212.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), Robust Estimates of Location, Princeton: Princeton University Press.
- Beran, Rudolf (1977a), "Robust Location Estimates," Annals of Statistics, 5, 431-444.
- _____ (1977b), "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics, 5, 445-463.
- Bickel, P. J., and Lehmann, E. L. (1975a), "Descriptive Statistics for Nonparametric Models, I. Introduction," Annals of Statistics, 3, 1038-1044.
- _____ (1975b), "Descriptive Statistics for Nonparametric Models, II. Location," Annals of Statistics, 3, 1045-1069.
- Bickel, P. J. (1976), "Another Look at Robustness: A Review of Reviews and Some New Developments," Scandinavian Journal of Statistics, 3, 145-168.
- Bölthausen, E. (1977), "Convergence in Distribution of Minimum-Distance Estimators," Metrika, 24, 215-227.
- Boos, Dennis D., and Serfling, R. J. (1977), "On Differentials, Asymptotic Normality and Almost Sure Behavior of Statistical

- Functions, With Application to M-statistics for Location Parameters," Technical Report No. M415, Department of Statistics, Florida State University.
- Easterling, Robert G. (1976), "Goodness of Fit and Parameter Estimation," Technometrics, 18, 1-9.
- Fisher, Ronald A. (1973), Statistical Methods and Scientific Inference, 3rd edition, New York: Hafner Press.
- Fu, James C. (1976), "A Note on Linear Estimators of Robust Estimation," (abstract) Institute of Mathematical Statistics Bulletin, 5, 123.
- Gross, A. M. (1973), "A Monte Carlo Swindle for Estimators of Location," Applied Statistics, 22, 347-353.
- Hampel, F. R. (1968), "Contributions to the Theory of Robust Estimation," Unpublished Ph.D. thesis, University of California, Berkeley.
- _____ (1971), "A General Qualitative Definition of Robustness," Annals of Mathematical Statistics, 42, 1887-1896.
- _____ (1974), "The Influence Curve and Its Role in Robust Estimation," Journal of the American Statistical Association, 69, 383-393.
- Huber, P. J. (1972), "Robust Statistics: A Review," Annals of Mathematical Statistics, 43, 1041-1067.
- _____ (1977), Robust Statistical Procedures, Philadelphia: Society for Industrial and Applied Mathematics.

- Jeffreys, Harold (1961), Theory of Probability, Oxford, Clarendon Press.
- Kempthorne, Oscar (1965), "The Classical Problem of Inference - Goodness of Fit," Proceedings of the Fifth Berkeley Symposium, 1, 235-249.
- Knusel, L. F. (1969), "Über Minimum-Distance-Schätzungen,"
Published Ph.D. thesis, Swiss Federal Institute of Technology.
- Littell, R. C. and Rao, P. V. (1975), "Estimation Based on the Kolmogorov-Smirnov Goodness of Fit Statistic," Technical Report 98, University of Florida.
- Parr, W. C. (1979), "A Conditional Property of Adaptive Estimators," submitted for publication.
- Parr, W. C. and DeWet (1979), "Minimum CVM-Norm Parameter Estimation," submitted for publication.
- Rao, P. V., Schuster, E. F., and Littell, R. C. (1975), "Estimation of Shift and Center of Symmetry Based on Kolmogorov-Smirnov Statistics," Annals of Statistics, 3, 862-873.
- Sahler, W. (1968), "A Survey on Distribution-Free Statistics Based on Distances Between Distribution Functions," Metrika, 13, 149-169.
- _____ (1970), "Estimation by Minimum-Discrepancy Methods," Metrika, 16, 85-106.

Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," Journal of the American Statistical Association, 69, 730-737.

Takeuchi, K. (1975) discussion of session on Robust Statistics, Bulletin of the International Statistical Institute, 46, 388.

Wolfowitz, J. (1957), "The Minimum Distance Method," Annals of Mathematical Statistics, 28, 75-88.