

ON MINIMUM CRAMER-VON MISES-NORM
PARAMETER ESTIMATION

by

W. C. Parr
T. De Wet

Technical Report No. 142
Department of Statistics ONR Contract

October 1980

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

ON MINIMUM CRAMER-VON MISES-NORM
PARAMETER ESTIMATION

W. C. Parr

Institute of Statistics, Texas A&M University
College Station, Texas 77843

T. De Wet

Department of Mathematical Statistics, Rhodes University
Grahamstown 6140, South Africa

ABSTRACT

Minimum distance parameter estimation using weighted Cramér-von Mises statistics is considered for the general one-dimensional case. Under rather general conditions, the derived estimators are asymptotically normal. Consideration is given to appropriate weights to produce Fisher-efficient estimators. In fact, estimators can be obtained with influence curves proportional to any desired smooth function, and hence prescribed first-order robustness properties. Many such curves (any "redescending" influence curve) are shown to require weight functions which take on negative values.

1. INTRODUCTION AND NOTATION

We consider minimum distance (MD) estimation utilizing the weighted Cramer-von Mises discrepancy. Letting $\Gamma = \{F_\theta, \theta \in \Omega\}$ denote a parametrized family of distribution functions (herein termed the model), and G_n denote the empirical distribution function based upon a random sample of size n from some distribution function G , we write

$$\delta_{\psi_\theta}(G_n, F_\theta) = \int_{-\infty}^{\infty} (G_n - F_\theta)^2 \psi_\theta f_\theta d\mu, \quad (1.1)$$

where μ denotes Lebesgue measure. The factor ψ_θ will be referred to as the "weight function", and will typically (although, as we shall see, not for a number of important cases of interest) be nonnegative and possess certain smoothness properties.

In the context of robust estimation, although we hope $G \in \Gamma$, i.e. $G = F_{\theta_0}$ for some $\theta_0 \in \Omega$, Γ is more realistically to be regarded as a model selected as containing a reasonable approximation to G . Even in the cases where $G \notin \Gamma$, minimization of $\delta_{\psi_\theta}(G_n, F_\theta)$ over all $\theta \in \Omega$ to obtain a MD-estimator typically (under broad regularity conditions) results in a situation where the estimand - that value for which the estimator is consistent - has an intrinsic probabilistic meaning in that it is associated with the best approximation (in a specified sense) to G in Γ . (Parr and Schucany (1980) give further discussion of this point.)

We shall formally define the MD-estimator based on G_n with respect to Γ and δ_{ψ_θ} as "the" solution of

$$\begin{aligned} \lambda_{G_n}(\theta) = \frac{\partial \delta_{\psi_\theta}(G_n, F_\theta)}{\partial \theta} &= -2 \int (G_n - F_\theta) \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta d\mu \\ &+ \int (G_n - F_\theta)^2 \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] d\mu = 0. \end{aligned} \quad (1.2)$$

Thus, $\lambda_{G_n}(T(G_n)) = 0$ defines our estimator $T(G_n)$, (we assume a unique method of choosing a consistent solution - by a local convexity argument, all consistent solutions will be $\sqrt{n}$ - equivalent.) and our estimand is a solution of $\lambda_G(T(G)) = 0$. In most of the typical contexts considered, another $\sqrt{n}$ - consistent estimator $\hat{\theta}$ of $T(G)$ will exist, and we may simply choose the solution of (1.2) closest to $\hat{\theta}$.

The influence curve (see Hampel (1974)) of the MD-estimator obtained as a root of (1.2), or equivalently by minimizing (1.1), may be seen by straightforward calculation to be (when $G = F_\theta$)

$$IC_{T, F_\theta}(c) = \frac{\int (\delta_c - F_\theta) \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta d\mu}{\int \left(\frac{\partial F_\theta}{\partial \theta} \right)^2 \psi_\theta f_\theta d\mu}, \quad (1.3)$$

$$-\infty < c < \infty (\delta_c(x) = I_{(c, \infty)}(x)), \text{ for all } \theta \in \Omega.$$

In Section 2 we examine efficient estimation for one-parameter problems, with the normal, double-exponential and t-distributions as examples, for the location problem, and the normal distribution for the scale problem. Extensions to multi-parameter situations are discussed. Section 3 consists of a series of comments on the robustness of estimators derived in this fashion, and exhibits asymptotic equivalents for some familiar "robust" estimators. Section 4 gives arguments for preferring MD estimators to other locally asymptotically equivalent, but perhaps computationally simpler, estimators. Section 5 discusses some suggestions regarding extension of the work in the preceding sections as well as its numerical implementation. While the discussion in these first five sections has been largely intuitive, avoiding technical details and regularity conditions, a proof of the main result is given in the appendix.

2. EFFICIENT ESTIMATION IN ONE-PARAMETER PROBLEMS

For many estimation problems there exists an expansion of the form

$$T[G_n] = T[F_\theta] + \frac{1}{n} \sum_{i=1}^n IC_{T, F_\theta}(X_i) + o_p\left(\frac{1}{\sqrt{n}}\right) \quad (2.1)$$

as $n \rightarrow \infty$.

This justifies (asymptotically) the usual interpretations of the influence curve, and tells us that $\sqrt{n}(T[G_n] - T[F_\theta]) \xrightarrow{d} N(0, E_{F_\theta}[IC_{T, F_\theta}^2(X)])$. For sufficiently "regular" estimators, a stronger expansion

$$T[H] = T[F_\theta] + \int IC_{T, F_\theta}(x) dH(x) + o(\|H - F_\theta\|) \quad (2.2)$$

$$\text{as } \|H - F_\theta\| \rightarrow 0$$

holds, where $\|\cdot\|$ is a norm on the space of distribution functions such that $\|G_n - G\| = o_p\left(\frac{1}{\sqrt{n}}\right)$ as $n \rightarrow \infty$, where G_n and G are as in Section 1 (see Huber (1977) or Boos and Serfling (1980)). From this, we can see that an estimator which is efficient in the Fisher-Rao sense will have

$$IC_{T, F_\theta}(c) = \frac{\frac{\partial \log f_\theta(c)}{\partial \theta}}{\int \left(\frac{\partial \log f_\theta}{\partial \theta} \right)^2 f_\theta d\mu} \quad (2.3)$$

Equating the right-hand sides of (1.3) and (2.3) and differentiating with respect to c , we find that when $IC_{T, F_\theta}(c)$ is continuous in c

$$\psi_{\theta}(c) \sim \frac{\frac{-\partial^2 \log f_{\theta}(c)}{\partial \theta \partial c}}{\frac{\partial F_{\theta}(c)}{f_{\theta}(c) \frac{\partial}{\partial \theta}}}, \quad (2.4)$$

yields in general $IC_{T, F_{\theta}}(c)$ proportional to the expression in (2.3), and in particular for location or scale models, the attained $IC_{T, F_{\theta}}(c)$ is exactly that of (2.3). Boos (1980) also gives (2.4) for the location case. Whether this holds for parameters of non-location or scale type must be determined by inspection in each case. Thus, except in special cases, full efficiency will not be consistent with the restriction, typical for goodness of fit applications, that $\psi_{\theta}(c) = \psi^*(F_{\theta}(c))$. Note that $\psi_{\theta}(\cdot)$ will vary with θ in the minimization.

Thus, minimization of $\delta_{\psi_{\theta}}(G_n, F_{\theta})$ with respect to θ should in many cases produce asymptotically fully efficient MD-estimators of θ , subject to the regularity conditions outlined in the appendix. For location estimation problems, (2.4) reduces to

$$\psi_{\theta}(c) = \frac{\frac{\partial^2 \log f_{\theta}(c)}{\partial c^2}}{f_{\theta}^2(c)}. \quad (2.5)$$

Hence, for the problem of estimation of the location parameter of a normal population (σ known and taken to be equal to 1 without loss of generality), the weight function for efficient estimation is found to be

$$\psi^*(c - \theta) = \psi_{\theta}(c) = 1/f^2(c - \theta) = 2\pi e^{-(c-\theta)^2}. \quad (2.6)$$

This yields $IC_{T, F_{\theta}}(c) = c - \theta$, $-\infty < c < \infty$ and an asymptotic

equivalent at the normal parent for the sample mean. Using the technique of De Wet (1980) it can be shown that the weighted Cramér-von Mises statistic with weight function (2.6) has the maximum approximate Bahadur slope in testing for the unit normal against shift alternatives. This "coincidence" is natural in the light of the results of Hodges and Lehmann (1963), with the new twist that in the present case the inverted test is not asymptotically normal.

The double exponential model provides an interesting case in which the ideal weight function is a point mass at 0, i.e. $\psi^*(x - \theta)$ is such that the integral

$$\int_A \psi^*(x - \theta) dx = I_A(0) \quad , \quad (2.7)$$

where $I_A(\cdot)$ is the indicator function. Note that this result does not follow from (2.5) but from the fact that the minimum resulting from use of (2.7) is precisely the sample median, when F_θ is strictly increasing at its median θ .

The t-distributions constitute a broad class ranging from the Cauchy to the normal. Here, the efficient weight function is given (for k degrees of freedom) by

$$\psi^*(x - \theta) = (k - (x - \theta)^2)(k + (x - \theta)^2)^{k-1} \quad . \quad (2.8)$$

Note that this weight function gives negative weight to extreme values of $x - \theta$, i.e. for $|x - \theta| > \sqrt{k}$. This corresponds in fact to the following basic principle. Since

$$\psi_\theta(x) = \frac{\partial IC_{T,F}(x)}{\partial x} / f_\theta^2(x) \quad (2.9)$$

is the weight function designed to result in an estimator with influence curve $IC_{T,F}(c)$ in the location problem, a redescending

influence curve can only be achieved by use of a weight function which is negative for some values. This may well serve to bolster the misgivings some feel regarding these redescenders - they are asymptotically equivalent to MD-estimators based on Cramér-von Mises discrepancies with weights which are not nonnegative!

Curiously, the optimal weight function for estimation of the standard deviation of a normal population with mean known and equal to 0 wlog turns out to be

$$\psi_{\theta}(x) = \frac{\theta^2}{f^2\left(\frac{x}{\theta}\right)}, \quad (2.10)$$

similar to the optimal weighting for location. This strongly suggests that simultaneous minimization of

$$\delta_{\psi_{\mu,\sigma}}(G_n, F_{\mu,\sigma}) = \int (G_n - F_{\mu,\sigma})^2 \psi_{\mu,\sigma}(x) f_{\mu,\sigma}(x) dx \quad (2.11)$$

with respect to μ and σ will yield efficient estimators of μ and σ in the normal location and scale problem, where

$$\psi_{\mu,\sigma}(x) = \sigma^2 / f^2\left(\frac{x - \mu}{\sigma}\right). \quad (2.12)$$

Note that we would actually define our estimators as the joint roots of the simultaneous equations

$$\frac{\partial \delta_{\psi_{\mu,\sigma}}(G_n, F_{\mu,\sigma})}{\partial \mu} = 0 \quad (2.13)$$

and

$$\frac{\partial \delta_{\psi_{\mu,\sigma}}(G_n, F_{\mu,\sigma})}{\partial \sigma} = 0.$$

A last intriguing example for location estimation is the logistic family (scale known), where it follows that the optimal weight function is that of the Anderson-Darling statistic,

$$\psi_{\theta}(x) = [F_{\theta}(x)(1 - F_{\theta}(x))]^{-1}.$$

For the simple exponential, $F(x, \theta) = 1 - e^{-x/\theta}$, $x \geq 0$, and the optimal weight is

$$\psi_{\theta}(x) = \frac{\theta^3}{x} e^{2x/\theta}. \quad (2.14)$$

3. ROBUST WEIGHTED CRAMÉR-VON MISES ESTIMATION

However, the ability to generate asymptotically efficient MD-estimators for a given parametric model would in itself be of little value, given the existence of asymptotically efficient L-estimators for location/scale problems (see Chernoff, Gastwirth and Johns (1967)) and the option of using maximum likelihood estimation for non-location/scale problems. Equation (1.3) gives the key to a more significant application, however. Differentiation of both sides of (1.3) with respect to c yields

$$\psi_{\theta}(c) = \frac{\frac{\partial IC_{T, F_{\theta}}(c)}{\partial c}}{f_{\theta}(c) \frac{\partial F_{\theta}(c)}{\partial \theta}} \quad (3.1)$$

as the weight function designed to give the associated estimators a specified (differentiable) influence curve $IC_{T, F_{\theta}}(\cdot)$. Our intuitive desire for a bounded weight function $\psi_{\theta}(\cdot)$ thus corresponds to requiring $IC_{T, F_{\theta}}(c)$ to be "extremely stable" for c in regions

where $f_{\theta}(c) \frac{\partial F_{\theta}(c)}{\partial \theta}$ is small (typically $c \rightarrow \pm \infty$).

Thus we may, as is the case with M-estimation, obtain estimators with influence curve proportional to any desired function (up to regularity conditions which will typically be satisfied for cases with $IC_{T, F_{\theta}}(\cdot)$ bounded and "smooth"). There is the further benefit that, opposed to M-estimation, when the model Γ is not true, but only contains a reasonable approximation to G , the value θ_0 for which the MD-estimator is consistent possesses an interpretable probabilistic meaning as discussed below in Section 4.

For the normal location problem, where the optimal weight function is

$$\psi_{\theta}(c) = 1/f^2(c - \theta) = O([u(1 - u)]^{-2} [\log(u(1 - u))]^{-1}), \quad (3.2)$$

as $c \rightarrow \pm \infty$, with $u = F(c - \theta)$, a natural modification to avoid the unbounded weight function is

$$\psi_{\theta}(c) = \begin{cases} 1/f^2(x - \theta) & |c - \theta| \leq k \\ 0 & |c - \theta| > k \end{cases} \quad (3.3)$$

for some fixed $k > 0$. This weight function "trims" the region of integration in the discrepancy $\delta_{\psi_{\theta}}(G_n, F_{\theta})$, and in fact yields a local almost sure $\sqrt{n}$ -equivalent of the trimmed mean with trimming proportion $1 - \Phi(k)$, where Φ is the unit normal cumulative. The local equivalence in fact suggests usage of the trimmed mean as a strategic starting value for the iterative procedure for computation of this MD-estimator.

Equation (3.1) suggests equivalents to other well-known "robust" estimators. Note, however, that the asymptotic $\sqrt{n}$ -equivalences hold in general only at the model - the estimators will have different behavior away from the strict parametric model. A local asymptotic equivalent to a general M-estimator of location $\hat{\theta}_n$ is provided (for any fixed density f) by taking

$$\psi_{\theta}(x) = \begin{cases} \frac{\partial \psi^* \left(\frac{x - \theta}{s} \right)}{\partial \theta} \\ f_{\theta}^2(x) \\ 0 \end{cases}, \quad \left| \frac{x - \theta}{s} \right| < \pi \quad (3.4)$$

otherwise ,

for some constant $s > 0$, where the M-estimator $\hat{\theta}_n$ is defined by

$$\frac{1}{n} \sum_{i=1}^n \psi^* \left(\frac{X_i - \hat{\theta}_n}{s} \right) = 0. \quad (3.5)$$

4. MINIMUM CVM NORM ESTIMATION AND OTHER METHODS

It may well be asked why one should use a minimum CVM norm estimator when other (computationally simpler) methods exist which are asymptotically equivalent when the model Γ is the correct one, that is, when $G \in \Gamma$. Section 5 deals briefly with the issue of computational simplicity. Reasons for preferring minimum CVM estimators include i) a concrete probabilistic interpretation for the estimator when $G \notin \Gamma$, a property not shared by the other methods; ii) the greater ease of applying minimum CVM norm estimators to complex problems not involving artificial symmetries, and their robustness properties; iii) desirable properties of minimum CVM norm estimators as indicated by Millar (1979); and iv) the extremely competitive small sample behavior of minimum CVM norm estimators as shown by Parr and Schucany (1980) in the location problem.

To understand the benefit of minimum CVM norm estimation as giving answers with concrete probabilistic interpretations, consider the case $G \notin \Gamma$. Here, under suitable regularity conditions, the estimator converges to the value $\theta_0 \in \Omega$ such that

$$\delta_{\psi_{\theta_0}}(G, F_{\theta_0}) = \inf_{\theta \in \Omega} \delta_{\psi_{\theta}}(G, F_{\theta}). \quad (4.1)$$

For instance, if $\psi_{\theta} = 1/f_{\theta}$, θ_0 gives a best L^2 approximate to G among the distribution functions $F_{\theta} \in \Gamma$. More generally, $\delta_{\psi_{\theta}}$ defines a notion of the "distance" measured in probability units between two distribution functions, and θ_0 minimizes that distance. Hence if $\delta_{\psi_{\theta}}$ is well chosen a reasonable approximation to G is obtained. Neither M or L estimators have this property. M-estimators converge (under suitable regularity conditions) to a solution of a linear equation in G having no necessary intrinsic meaning when $G \notin \Gamma$.

In many complex problems, the CVM based estimators are easier to apply than M or L estimators. The relative scarcity of L estimators proposed for non-location or scale problems which are robust and interpretable when $G \notin \Gamma$ serves to illustrate this point. Application of M-estimation to such problems involves the extremely complicated solution of the equations of Huber (1977, p.33). For nonsymmetric or non-additive errors, this is an unsolved problem for practical applications. By way of contrast, robustness against gross errors of a CVM based estimator can usually be achieved by keeping ψ_{θ} small (that is, $\int |\psi_{\theta}| f_{\theta} d\mu \leq c < \infty$ for all $\theta \in \Omega$), while still preserving consistency of the derived estimator when $G \in \Gamma$.

Millar (1979) indicates a number of desirable features of minimum CVM norm estimators in a precisely specified mini-max sense against sequences of alternatives approaching the model Γ . This further bolsters the intuitive notion that the estimators should have good properties when the model is not exactly true, but still contains a reasonable approximation to G .

Parr and Schucany (1980) report partial results from an extensive Monte Carlo study comparing minimum CVM norm estimators to the best of the M and L estimators for location examined in

the Princeton Study (Andrews, *et al.* (1972)). They find the CVM based estimators to be highly competitive with these others, which were chosen for inclusion because of their previously documented excellent behavior at the distributions examined in this later comparative study.

5. COMPUTATIONAL MATTERS, EXTENSIONS AND CONCLUSIONS

Unfortunately, for a variety of possible weight functions, the discrepancy $\delta_{\psi_{\theta}}(G_n, F_{\theta})$ does not admit a simple calculating form. However, the alternative version

$$\delta_{\psi_{\theta}}^*(G_n, F_{\theta}) = \frac{1}{n} \sum_{i=1}^n (F_{\theta}(X_{(i)}) - \frac{i}{n+1})^2 \psi_{\theta}(F_{\theta}^{-1} \frac{i}{n+1}) \quad (5.1)$$

(where $X_{(i)}$ is the i^{th} order statistic) which yields an estimator with the same asymptotic distribution as the one obtained by inverting $\delta_{\psi_{\theta}}(G_n, F_{\theta})$, is much easier to calculate when $\delta_{\psi_{\theta}}(G_n, F_{\theta})$ does not integrate in closed form. In fact, the derivatives of $\delta_{\psi_{\theta}}^*(G_n, F_{\theta})$ generally possess simple forms, permitting the use of a Newton-Raphson routine for computation of the estimator. Thus, the estimators are essentially no more costly to compute than the usual M-estimators. (See DeWet and Venter (1973) where this statistic is used in a goodness-of-fit setting).

Alternatively, in the location case, where for reasons of invariance $\psi_{\theta}(F_{\theta}^{-1}\{\frac{i}{n+1}\}) = h(\frac{i}{n+1})$, a function free of θ , weighted nonlinear least squares techniques can be employed to minimize $\delta_{\psi_{\theta}}^*(G_n, F_{\theta})$. In experimentation involving this second method, the first author has used standard nonlinear regression packages to compute the MD-estimator and obtained convergence to sufficient accuracy for applications typically in less than five iterations.

Although (2.10) gave a single weight function which yielded jointly optimal estimators for the location and scale of a normal distribution, a single such weight function does not exist for multiparameter parameter problems in general. For a p dimensional parameter $\underline{\theta}' = (\theta_1, \theta_2, \dots, \theta_p)$, jointly optimal estimators are however available as the joint roots in $\theta_1, \theta_2, \dots, \theta_p$ of

$$\frac{\partial \delta_{\psi_{\underline{\theta},i}}(G_n, F_{\underline{\theta}})}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, p \quad (5.2)$$

where

$$\psi_{\underline{\theta},i} = \frac{\frac{-\partial^2 \log f_{\underline{\theta}}(c)}{\partial \theta_i \partial c}}{\frac{\partial F_{\underline{\theta}}(c)}{\partial \theta_i} - \frac{f_{\underline{\theta}}(c)}{f_{\underline{\theta}}(c)}}, \quad i = 1, \dots, p. \quad (5.3)$$

Such a procedure would be necessary if optimality of the vector estimator $\hat{\underline{\theta}}$ were important. It will be noted that equations (5.2) correspond to (in fact under general conditions are asymptotically equivalent to) the likelihood equations.

6. APPENDIX

In this section we give conditions for the MD-estimator based upon $\delta_{\psi_{\underline{\theta}}}(G_n, F_{\underline{\theta}})$ to be asymptotically normal (and Fisher-Rao efficient with proper choice of $\psi_{\underline{\theta}}$, although our result allows other weightings). We assume that strong consistency of the selected root of (1.2) is guaranteed by some result such as Theorem 2 of Parr and Schucany (1980), or one of its modifications stated there, we take $\theta_o = 0$ without loss of generality, write $F_{\theta_o} = F = G$, $f_{\theta_o} = f$, and define

$$h(t) = \begin{cases} \frac{\lambda_F(t)}{t} & t \neq 0 \\ \lambda'_F(T(F)) & t = 0 \end{cases} \quad (6.1)$$

(Note $\lambda_F(t(F)) = T(F) = 0$, and hence $h(t)$ is just the difference quotient when $t \neq T[F]$. Thus, $h(t)$ is continuous at $t = T[F]$ by differentiability of $\lambda_F(c)$ at $T[F]$).

It is sufficient for asymptotic normality that $IC_{T,F}(c)$ as in (1.3) be square integrable with respect to F , and

$$T[G_n] - T[G] - H(G_n) \int IC_{T,F}(c) dG_n(c) = o_p\left(\frac{1}{\sqrt{n}}\right), \quad (6.2)$$

where $H(G_n) \rightarrow 1$ w.p.1. We take

$$H(G_n) = \frac{\lambda'_G(T[G])}{h(T[G_n])} \quad (6.3)$$

and then

$$\begin{aligned} & |T[G_n] - T[F] - H(G_n) \int IC_{T,F}(c) dG_n(c)| = \\ & \left| \frac{1}{h(T[G_n])} \right| \left| 2 \int (G_n - F) \left\{ \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=T[G_n]} - \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=0} \right\} d\mu \right. \\ & \left. + \int (F - G_n) (F + G_n - 2F_\theta) \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] \Big|_{\theta=T[G_n]} d\mu \right| \quad (6.4) \end{aligned}$$

after some elementary algebraic operations. Since $h(T[G_n]) \rightarrow \lambda'_F(T(F)) > 0$ w.p.1 by assumption, we need merely show each of the terms in the second factor to be $o_p\left(\frac{1}{\sqrt{n}}\right)$. For the first

term it is clear that

$$\left| 2 \int (G_n - F) \left\{ \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=T[G_n]} - \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=0} \right\} d\mu \right| \leq$$

$$2 \sup \frac{|G_n - F|}{\{F(1 - F)\}^{1/2 - \varepsilon}} \int \{F(1 - F)\}^{1/2 - \varepsilon} \cdot \left| \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=T[G_n]} - \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=0} \right| d\mu. \quad (6.5)$$

Since $\sup \frac{|G_n - F|}{\{F(1 - F)\}^{1/2 - \varepsilon}} = o_p\left(\frac{1}{\sqrt{n}}\right)$ for $0 < \varepsilon < 1/2$ and $T[G_n]$

is strongly consistent, it follows that a sufficient condition for the first term to be of proper order is

$$\int \{F(1 - F)\}^{1/2 - \varepsilon} \left| \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=c} - \frac{\partial F_\theta}{\partial \theta} \psi_\theta f_\theta \Big|_{\theta=0} \right| d\mu \rightarrow 0, \quad (6.6)$$

as $c \rightarrow 0$. (For specific cases such as location estimation, this simplifies considerably).

For the second term in the sum on the RHS of (6.4), it suffices to show that both

$$\int (F - G_n)^2 \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] \Big|_{\theta=T[G_n]} d\mu = o_p\left(\frac{1}{\sqrt{n}}\right) \quad (6.7)$$

and

$$2 \int (F - G_n)(G - F_\theta) \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] \Big|_{\theta=T[G_n]} d\mu = o_p\left(\frac{1}{\sqrt{n}}\right). \quad (6.8)$$

The first of these [viz (6.7)] is bounded (for $0 < \delta < 1$) in absolute value by

$$\sup \left(\frac{|G_n - F|^2}{\{F(1 - F)\}^{1-\delta}} \cdot \int \{F(1 - F)\}^{1-\delta} \left| \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] \Big|_{\theta=T[G_n]} \right| d\mu \right) \quad (6.9)$$

and the second (6.8) (for $0 < \varepsilon < 1/2$) by

$$2 \sup \left(\frac{|G_n - F|}{\{F(1 - F)\}^{1/2 - \varepsilon}} \cdot \int \{F(1 - F)\}^{1/2 - \varepsilon} \left| (F - F_\theta) \frac{\partial}{\partial \theta} [\psi_\theta f_\theta] \Big|_{\theta=T[G]} \right| d\mu \right) \quad (6.10)$$

Thus if

$$\int \{F(1 - F)\}^{1-\delta} \frac{\partial}{\partial \theta} [\psi_{\theta} f_{\theta}] d\mu < \infty$$

for θ in some neighborhood of 0, and

$$\lim_{\theta \rightarrow 0} \int \{F(1 - F)\}^{1/2 - \varepsilon} (F - F_{\theta}) \frac{\partial}{\partial \theta} [\psi_{\theta} f_{\theta}] d\mu = 0,$$

these terms are also $O_p\left(\frac{1}{\sqrt{n}}\right)$. Collecting our conditions, with all notation as before, we thus have the following.

Theorem. If $\{T[G_n]\}_{n=1}^{\infty}$ is a sequence of D-estimators based upon the sequence $\{G_n\}_{n=1}^{\infty}$, with respect to Γ and $\delta_{\psi_{\theta}}$, chosen as a solution of (1.2), and

- i) $T[G_n]$ is strongly consistent,
- ii) $G = F_{\theta}$ for some $\theta \in \Omega$ (taken without loss of generality to be $\theta = 0$),
- iii) $\lambda_F(c)$ is differentiable at $T[F]$ and $\lambda'_F(T(F)) > 0$,
- iv) $0 < \int IC_{T,F}^2(c) dF(c) = \sigma^2 < \infty$, and

$$\begin{aligned} \text{v) } \lim_{c \rightarrow 0} \int \{F(1 - F)\}^{1/2 - \varepsilon} \left| \frac{\partial F}{\partial c} \psi_c f_c - \frac{\partial F}{\partial \theta} \psi_{\theta} f_{\theta} \right|_{\theta=0} d\mu &= 0, \\ \int \{F(1 - F)\}^{1-\varepsilon} \frac{\partial}{\partial \theta} [\psi_{\theta} f_{\theta}] d\mu &< \infty \text{ for } \theta \text{ in some neighborhood of } 0, \end{aligned}$$

and $\lim_{\theta \rightarrow 0} \int \{F(1 - F)\}^{1/2 - \varepsilon} \left| (F - F_{\theta}) \frac{\partial}{\partial \theta} [\psi_{\theta} f_{\theta}] \right| d\mu = 0$ for some $0 < \varepsilon < 1/2$, then $\sqrt{n}[T[G_n] - T[G]] \xrightarrow{d} N(0, \sigma^2)$. (In fact, under these conditions we have an almost sure representation of the estimator, as can be easily seen from an examination of the proof).

It should be noted that finiteness of $\int \frac{\partial F}{\partial \theta} \psi_{\theta} f_{\theta} d\mu$ and $\int \frac{\partial}{\partial \theta} [\psi_{\theta} f_{\theta}] d\mu$ for θ in a neighborhood of zero is sufficient but not necessary for iii)-v) to be satisfied. Thus, for the case of location estimation, a set of sufficient conditions to replace iii)-v) is

iii*) ψ_0 and $\left. \frac{\partial \psi_\theta}{\partial \theta} \right|_{\theta=0}$ are bounded

iv*) $\int f_0^2 d\mu < \infty$

and v*) $\int f_0' d\mu < \infty$.

These conditions are quite reasonable due to robustness considerations which make bounded weights ψ_θ desirable and the requirement of a square integrable density with finite Fisher's information a natural smoothness requirement. Boos (1980) gives moment-type conditions for the analogue of the above theorem when θ is a location parameter.

7. ACKNOWLEDGEMENTS

The research of the first author was partially supported by ONR Contract N00014-75-C-0439, and that of the second author by the C.S.I.R. The authors are grateful to W. R. Schucany and to the referees for a number of comments which improved the presentation and content of the manuscript.

BIBLIOGRAPHY

- 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 University Press, Princeton, New Jersey.
- Boos, D. D. (1980). Minimum distance estimators for location and goodness-of-fit. To appear in *Journal of the American Statistical Association*.
- Boos, D. D. and Serfling, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions with application to M-estimates, *Ann. Statist.*, 8, 618-624.
- Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* 38, 52-72.

- De Wet, T. (1980). Cramér-von Mises tests for independence. *J. Multivar. Anal.*, 10, 38-50.
- De Wet, T. and Venter, J. H. (1973). Asymptotic distribution for quadratic forms with applications to tests of fit. *Ann. Statist.* 1, 380-387.
- Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* 5, 110-123.
- Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* 42, 1887-1896.
- Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* 69, 383-393.
- Hodges, J. L. and Lehmann, E. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* 34, 598-611.
- Huber, P. J. (1977). *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics. Bristol, England.
- Millar, P. W. (1979). Robust estimation via minimum distance methods. Unpublished manuscript.
- Parr, W. C. and Schucany, W. R. (1980). Robust estimation and minimum distance. *Journal of the American Statistical Association*, 75, 616-624.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 142	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Minimum Cramer-Von Mises-Norm Parameter Estimation		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER 142
7. AUTHOR(s) W. C. Parr T. De Wet		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 280
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Va. 22217		12. REPORT DATE October 1980
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Minimum distance parameter estimation using weighted Cramer-von Mises statistics is considered for the general one-dimensional case. Under rather general conditions, the derived estimators are asymptotically normal. Consideration is given to appropriate weights to product Fisher-efficient estimators. In fact, estimators can be obtained with influence curves proportional to any desired smooth function, and hence prescribed first-order robustness properties. Many such curves (any "redescending" influence curve) are shown to require weight functions which take on negative values.		