The Canadian Journal of Statistics
Vol. 19, No. 2, 1991, Pages 179–190
La Revue Canadienne de Statistique

Distributions minimizing Fisher information for scale in Kolmogorov neighbourhoods

Douglas P. WIENS* and K.H. Eden WU†

University of Alberta and *The Chinese University of Hong Kong*

Key words and phrases: Robust estimation of scale, minimum Fisher information for scale, M-estimates of scale, minimax variance, Kolmogorov neighbourhood.
AMS 1985 subject classifications: Primary 62F35; secondary 62G05.

ABSTRACT

We construct those distributions minimizing Fisher information for scale in Kolmogorov neighbourhoods $\mathcal{K}_\epsilon(G) = \{F \mid \sup_x |F(x) - G(x)| \leq \epsilon\}$ of d.f.'s G satisfying certain mild conditions. The theory is sufficiently general to include those cases in which G is normal, Laplace, logistic, Student's t, etc. As well, we consider $G(x) = 1 - e^{-x}$, $x \geq 0$, and correct some errors in the literature concerning this case.

RÉSUMÉ

On construit les lois qui minimisent l'information de Fisher pour l'échelle dans des voisinages de Kolmogorov $\mathcal{K}_\epsilon(G) = \{F \mid \sup_x |F(x) - G(x)| \leq \epsilon\}$ de fonctions de répartition G, satisfaisant des conditions peu restrictives. La théorie est assez générale pour permettre à la loi G d'être, par exemple, gaussienne, de Laplace, logistique ou t de Student. On considère aussi le cas où $G(x) = 1 - e^{-x}$, $x \geq 0$, et certaines erreurs apparaissant dans la littérature à propos de celui-ci sont corrigées.

1. INTRODUCTION AND SUMMARY

In the theory of robust, minimax variance estimation as developed by Huber (1964, 1981), a frequently occurring problem is that of determining that member of a certain class of distributions, representing all "reasonable" departures from a "target" distribution, which minimizes the Fisher information for the quantity being estimated. Such departures are often modelled by Kolmogorov neighbourhoods:

$$\mathcal{K}_\epsilon(G) = \left\{ F \,\middle|\, \sup_{-\infty<x<\infty} |F(x) - G(x)| \leq \epsilon \right\}$$

in which ϵ and G are known and fixed.

Huber (1964) minimized the information for *location* in $\mathcal{K}_\epsilon(\Phi)$, $\epsilon \leq 0.0303$. Here, Φ is the standard normal d.f. Sacks and Ylvisaker (1972) extended this to the range $0.0303 \leq \epsilon \leq 0.5$. Wiens (1986) considered this problem for general, symmetric G. Collins and Wiens (1989) extended these results to Lévy neighbourhoods of d.f.'s G satisfying conditions similar to those imposed in Wiens (1986).

*Research supported by a Natural Sciences and Engineering Research Council of Canada grant.
† Research carried out while a graduate student at the University of Alberta.

The problem of minimizing the information for *scale* in Kolmogorov neighbourhoods has hitherto not received a systematic treatment in the literature. Note that if σ is a scale parameter for a r.v. X, then $\log \sigma$ is a location parameter for $\log |X|$. By this device, certain results for location estimation may be transferred to the problems of scale estimation. This approach was taken by Huber (1981) in minimizing the information for scale in the gross-errors neighbourhood

$$\mathcal{G}_\epsilon(\Phi) = \{F = (1-\epsilon)\Phi + \epsilon H,\ H \text{ arbitrary}\}.$$

The log transformation was useful here, since under it the gross-errors structure is maintained *and* there already existed a theory of minimum information for location for $\mathcal{G}_\epsilon(G)$, with G nonsymmetric.

The Kolmogorov neighbourhood structure is maintained under the log transformation. However, a common requirement in all of the aforementioned papers on minimizing information for location in $\mathcal{K}_\epsilon(G)$ is that G be symmetric. This symmetry is typically destroyed by taking the logarithms. Thus, the problem of minimizing the information for scale requires either a direct approach, or the derivation of a location estimation theory which is valid in $\mathcal{K}_\epsilon(G)$ for nonsymmetric G.

Each approach has its merits. The former was used in the Ph.D. thesis of Wu (1990), upon which the present paper is largely based. For added clarity of presentation we now adopt the latter approach. This allows us to present some preliminary results as extensions of location-theory results of Huber (1964, 1981) and Wiens (1986).

We begin with some preliminary definitions and motivating remarks, then outline the transformation to a location problem and present some general theory. In Section 3 some specific solutions are presented. These are general enough to include the cases $G = \Phi, G$ the Laplace d.f., and more generally $G = G_l$ with density proportional to $\exp(-|x|^l/l), l > 0$. The logistic and Student's t distributions are covered as well. We also consider the case in which $G(x) = 1 - e^{-x}, x \geq 0$; in so doing we correct some errors made in Thall (1979).

2. GENERAL THEORY

The proof of Theorem 1 below is completely analogous to that of Theorem 4.4.2 of Huber (1981), and so is omitted.

DEFINITION . *Fisher information for scale of a distribution F on the real line is*

$$I(F;\,1) = \sup_{\chi} \frac{\left\{\int_{-\infty}^{\infty} x\chi'(x)\,dF(x)\right\}^2}{\int_{-\infty}^{\infty} \chi^2(x)\,dF(x)},$$

where the sup is taken over all continuously differentiable functions χ with compact support, satisfying $\int_{-\infty}^{\infty} \chi^2(x)\,dF(x) > 0$.

THEOREM 1. *The following two assertions are equivalent:*

(1) $I(F;\,1) < \infty$.

(2) *F has a density f, absolutely continuous on $\mathbb{R} \setminus \{0\}$, satisfying:*

(i) $xf(x) \to 0$ *as* $x \to 0, \pm\infty$;
(ii) $\int_{-\infty}^{\infty} (-x\frac{f'}{f}(x) - 1)^2 f(x)\,dx < \infty$.

In either case, we have

$$I(F;\,1) = \int \left(-x\frac{f'}{f}\,(x) - 1\right)^2 f(x)\,dx < \infty.$$

REMARK 1. Define $F_\sigma(x) = F(x/\sigma)$ for $\sigma > 0$. Then if $I(F;\,\sigma)$ is defined as $I(F;\,1)$, we have that the value of this functional is $(1/\sigma^2)I(F;\,1)$.

REMARK 2. An M-estimate of scale is defined as $S(F_n)$, where F_n is the empirical distribution function based on a sample $X_1, \ldots, X_n \sim F$, and the functional $S(F)$ is defined implicitly by

$$\int_{-\infty}^{\infty} \chi\left(\frac{x}{S(F)}\right) dF(x) = 0 \tag{2.1}$$

Under appropriate regularity conditions [see for example Boos and Serfling (1980) and Serfling (1981)]

$$\sqrt{n}\,\{\log S(F_n) - \log S(F)\} \xrightarrow{w} \mathsf{N}(0,\ V(\chi,\ F)), \tag{2.2}$$

where

$$V(\chi,\ F) = \frac{\displaystyle\int_{-\infty}^{\infty} \chi^2\left(\frac{x}{S(F)}\right) dF(x)}{\left\{\displaystyle\int_{-\infty}^{\infty} \chi'\left(\frac{x}{S(F)}\right)\frac{x}{S(F)}\,dF(x)\right\}^2}.$$

Now let $\mathcal{F}$ be a given convex class of distributions, and suppose that F_0 minimizes $I(F;\,1)$ in $\mathcal{F}$. Define

$$\chi_0(x) = -\frac{f_0{}'}{f_0}\,(x) - 1,$$

corresponding to maximum-likelihood estimation of σ if $X_1, \ldots, X_n \sim F_{0,\sigma}$. Define $S_0(F)$ by (2.1), with $\chi = \chi_0$. Then $S_0(F_0) = 1$. We have

$$V(\chi_0,\ F) \le V(\chi_0,\ F_0) = \frac{1}{I(F_0;\,1)} \le V(\chi,\ F_0) \tag{2.3}$$

for all $F \in \mathcal{F}_1 = \{F \in \mathcal{F} \mid S_0(F) = 1\}$ and all χ such that (2.1) holds for $F \in \mathcal{F}_1$. The second inequality in (2.3) is essentially the Cramér-Rao inequality; the first is established by variational arguments, as in Huber (1964, 1981). If follows from (2.3) that

$$\sup_{\mathcal{F}_1} V(\chi_0,\ F) = \inf_{\chi} \sup_{\mathcal{F}_1} V(\chi,\ F),$$

so that χ_0 yields a minimax variance estimate of scale for $F \in \mathcal{F}_1$.

The question of whether or not the saddle-point property (2.3) extends to all of $\mathcal{F}$ is considered, for $\mathcal{F} = \mathcal{K}_\epsilon(G)$ and $\mathcal{F} = \mathcal{G}_\epsilon(G)$, in Wiens and Wu (1990). We note that this question was answered, in the affirmative, by Huber (1981) in the class $\mathcal{G}_\epsilon(\Phi), \epsilon \le 0.04$. For $\mathcal{K}_\epsilon(G)$ the answer is negative, for all $\epsilon > 0$. Note that the failure of the saddlepoint property does not imply that χ_0 fails to yield a minimax estimate of scale for all of $\mathcal{F}$. To verify the minimax property, however, without the benefit of (2.3), would require one to perform the, evidently very intractable, task of determining $\sup_{\mathcal{F}} V(\chi,\ F)$ for each fixed χ, and to then determine the infimum thereof.

Suppose that a random variable X has d.f. $F \in \mathcal{K}_\epsilon(G)$, and that G is symmetric. Let F_- be the d.f. of $-X$, and put $\bar{F} = \frac{1}{2}F + \frac{1}{2}F_-$. Then $\bar{F} \in \mathcal{K}_\epsilon(G), \bar{F}$ is symmetric, and, since $I(F;\,1)$ is a convex functional of F,

$$I(\bar{F};\,1) \le I(F;\,1) = I(F_-;\,1).$$

We may thus restrict to symmetric $F \in \mathcal{K}_\epsilon(G)$. Then if as well $I(F;\,1) < \infty, Y = \log|X|$ has d.f. and density

$$\begin{aligned} \tilde{F}(y) &= F(e^y) - F(-e^y) = 2F(e^y) - 1, \\ \tilde{f}(y) &= 2e^y f(e^y), \qquad -\infty < y < \infty. \end{aligned} \tag{2.4}$$

We have $F \in \mathcal{K}_\epsilon(G)$ iff $\tilde{F} \in \mathcal{K}_{2\epsilon}(\tilde{G})$. Furthermore, Fisher information for scale, of F, translates into Fisher information for location, of $\tilde{F}$. By virtue of the location-theory analogue of Theorem 1, we may define the latter via

$$I_*(F) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(-\frac{f'}{f}(y)\right)^2 f(y)\,dy & \text{if } F \text{ has an absolutely continuous density } f; \\ \infty & \text{otherwise.} \end{cases}$$

Then $I_*(\tilde{F}) = I(F;\,1)$, and so we seek to minimize $I_*(\tilde{F})$ over $\mathcal{K}_{2\epsilon}(\tilde{G}), 0 \le \epsilon \le 0.25$.

We will assume that $I_*(\tilde{G}) < \infty$; it then follows from Vandelinde (1979) that the finite information members are weakly dense in $\mathcal{K}_{2\epsilon}(\tilde{G})$. As in Huber (1964), we may then conclude that there exists an information-minimizing $\tilde{F}_0$.

Necessary and sufficient properties of $\tilde{F}_0$ may now be obtained by a straightforward extension of the theory in Sections 1 and 2 of Wiens (1986). First define

$$\psi_0(y) = -\frac{\tilde{f}_0{}'}{\tilde{f}_0}(y),$$

and note that

$$\chi_0(x) = \psi_0(\log|x|). \tag{2.5}$$

Since $I_*(\tilde{F})$ is convex, $I_*(\tilde{F}_0) = \min I_*(\tilde{F})$ iff, for each $\tilde{F} \in \mathcal{K}_{2\epsilon}(\tilde{G})$ with $I_*(\tilde{F}) < \infty$,

$$\begin{aligned} 0 &\le \frac{d}{dt} I_*\big((1-t)\tilde{F}_0 + t\tilde{F}_1\big)\big|_{t=0} \\ &= \int_{-\infty}^{\infty} \{2(\tilde{f}_0 - \tilde{f})'\psi_0 + (\tilde{f}_0 - \tilde{f})\psi_0^2\}\,dy. \end{aligned} \tag{2.6}$$

This condition becomes more useful if an integration by parts is possible. Define an operator J on the class of absolutely continuous functions on $\mathbb{R}$ by

$$J(\psi)(y) = 2\psi'(y) - \psi^2(y).$$

Extend J by left continuity where ψ' is discontinuous. Note that, if we may integrate by parts,

$$\int J(\psi_0)\,d\tilde{F}_0 = I_*(\tilde{F}_0). \tag{2.7}$$

LEMMA 2. *If $\tilde{F}_0$ is such that ψ_0 is absolutely continuous and bounded, then in order that $\tilde{F}_0$ minimize $I_*(\tilde{F})$ in $\mathcal{K}_{2\epsilon}(\tilde{G})$ it is necessary and sufficient that*

(1) $\tilde{F}_0 \in \mathcal{K}_{2\epsilon}(\tilde{G})$;
(2) $0 \leq \int_{-\infty}^{\infty} J(\psi_0)(x)\, d(\tilde{F}_1 - \tilde{F}_0)(x)$ *for all* $\tilde{F}_1 \in \mathcal{K}_{2\epsilon}(\tilde{G})$ *with* $I_*(\tilde{F}_1) < \infty$.

Proof. Integrate by parts in (2.6), using condition (2)(i) of Theorem 1. Q.E.D.

Motivated by the preceding discussion, we may now formally state the assumptions on G:

(G1) G is symmetric and strictly increasing on $(-\infty, \infty)$.
(G2) $0 < I(G;\, 1) < \infty$.
(G3) The function $\tilde{\xi}(y) = -\tilde{g}'(y)/\tilde{g}(y)$, defined throughout $\mathbb{R}$ by virtue of (G1) and (G2), is twice continuously differentiable.
(G4) There is a point M such that $J(\tilde{\xi})(y)$ is strictly increasing for $y < M$, it is strictly decreasing for $y > M$, and $J(\tilde{\xi})(M) > 0$.

REMARK 3. Assumptions (G1)–(G4) hold for the logistic distribution, all Student's t-distributions, and those G_l defined in Section 1, for $l > 0$.

The proof of the following lemma is very similar to that of Lemma 1 of Wiens (1986), and so is omitted.

LEMMA 3. *Under assumptions* (G1)–(G4), $\tilde{\xi}(y)$ *is strictly increasing on* $\mathbb{R}$.

In Section 3 below we exhibit the minimum-information members $\tilde{F}_0$ of $\mathcal{K}_{2\epsilon}(\tilde{G})$ for a variety of distributions G satisfying (G1)–(G4). In each case, that these $\tilde{F}_0$ do minimize $I_*(\tilde{F})$ following from the following theorem.

THEOREM 4. *If $\tilde{F}_0$ possesses the following properties, then it is the unique member of $\mathcal{K}_{2\epsilon}(\tilde{G})$ minimizing $I_*(\tilde{F})$ over $\mathcal{K}_{2\epsilon}(\tilde{G})$:*

(S1) $\tilde{F}_0 \in \mathcal{K}_{2\epsilon}(\tilde{G})$.

(S2) ψ_0 *is absolutely continuous.*

(S3) *There exists a sequence* $-\infty = a_1 < b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_{n-1} < b_{n-1} \leq a_n < b_n = \infty$, *and constants* $\lambda_1, \ldots, \lambda_n$ *with* $\lambda_1, \lambda_n < 0$, *such that:*

(i) $$J(\psi_0)(y) = \begin{cases} \lambda_i, & a_i < y \leq b_i,\ i = 1, \ldots, n, \\ J(\tilde{\xi})(y), & b_i < y \leq a_{i+1},\ i = 1, \ldots, n-1. \end{cases}$$
(ii) *With* $B_L := \{y \mid \tilde{F}_0(y) = \tilde{G}(y) - 2\epsilon\}$ *and* $B_U := \{y \mid \tilde{F}_0(y) = \tilde{G}(y) + 2\epsilon\}$,

$$B_U \cup B_L = \bigcup_{i=1}^{n-1} [b_i, a_{i+1}].$$

Furthermore, each $[b_i, a_{i+1}]$ *is contained in exactly one of* B_L, B_U.
(iii) *If* $a_i \in B_L\,[B_U]$ *then* $J(\psi_0)(a_i^+) = \lambda_i \leq [\geq] J(\psi_0)(a_i^-)$. *If* $b_i \in B_L\,[B_U]$ *then* $J(\psi_0)(b_i^-) = \lambda_i \geq [\leq] J(\psi_0)(b_i^+)$.
(iv) *If* (b_i, a_{i+1}) *is nonempty and contained in* $B_L\,[B_U]$, *then* $J(\tilde{\xi})$ *is weakly decreasing [increasing] there.*

Proof. The possible solutions to the equation $J(\psi_0)(y) =$ constant are given in Remark 4

below. By (S3)(i) we may infer that

$$\psi_0(y) = -\sqrt{|\lambda_1|}, \quad \tilde{f}_0(y) \propto e^{-\sqrt{|\lambda_1 y|}} \qquad \text{for} \quad y < b_1,$$

$$\psi_0(y) = \sqrt{|\lambda_n|}, \quad \tilde{f}_0(y) \propto e^{-\sqrt{|\lambda_n y|}} \qquad \text{for} \quad y > a_n.$$

This is because the $\tilde{f}_0$'s corresponding to the "tanh" or "coth" forms of ψ_0 are not integrable on half-infinite intervals.

In particular, ψ_0 is bounded. Conditions (S1) and (S2) now allow us to apply Lemma 2, and to satisfy condition (1) of the lemma. Condition (S3) guarantees that condition (2) of Lemma 2 is satisfied. To see this, split the range of integration up into intervals and then integrate by parts over those nonempty intervals (b_i, a_{i+1}), using (G3). The integral can then be rearranged as a sum of nonnegative terms, using (iii) and (iv) of (S3).

The uniqueness of $\tilde{F}_0$ now follows from Proposition 4.4.5 of Huber (1981), if

(i) $\tilde{f}_0(x) > 0$, $x \in (-\infty, \infty)$;
(ii) $0 < I_*(\tilde{F}_0) = \int_{-\infty}^{\infty} \psi_0^2(x)\, d\tilde{F}_0(x)$.

These follow from (G1), (S2), and the observation that no solution to the equation $J(\psi_0) =$ constant can remain bounded as $\tilde{f}_0(x) \longrightarrow 0$. Q.E.D.

REMARK 4. The possible solutions to the equation $J(\psi_0)(y) =$ constant are given by:

(i) $J(\psi_0)(y) = \lambda^2$:

$$\psi_0(y) = \lambda \tan\left(\frac{\lambda}{2}(y - w)\right), \qquad \tilde{f}_0(y) \propto \cos^2\left(\frac{\lambda}{2}(y - w)\right);$$

(ii) $J(\psi_0)(y) = -\lambda^2$:

$$\psi_0(y) = \lambda, \ \lambda \tanh\left(-\frac{\lambda}{2}(y - w)\right), \text{ or } \lambda \coth\left(-\frac{\lambda}{2}(y - w)\right);$$

correspondingly

$$\tilde{f}_0(y) \propto e^{-\lambda y}, \ \cosh^2\left(-\frac{\lambda}{2}(y - w)\right), \text{ or } \sinh^2\left(-\frac{\lambda}{2}(y - w)\right).$$

REMARK 5. Condition (S1) forces the additional conditions

(S4)(i) $\tilde{f}_0(y) = \tilde{g}(y), y \in B_L \cup B_U$;
(S4)(ii) $\psi_0(y) - \tilde{\xi}(y) \leq 0$ on $B_L (\geq 0$ on $B_U)$.

Condition S4(ii) fails to follow, with equality, from S4(i) only when $[b_i, a_{i+1}]$ is a single point, at which $\tilde{F}_0$ is on one of the boundaries $\tilde{G} \pm 2\epsilon$. In this case (S4)(ii) follows from the identity

$$(\psi_0 - \tilde{\xi})(y) = \frac{d}{dy} \log \frac{\tilde{g}}{\tilde{f}_0}(y).$$

3. SOME CLASSES OF SOLUTIONS

The conditions of Section 2 fall short of determining $\tilde{F}_0$ completely. The solutions presented in this section, under assumptions (G1)–(G4), were obtained largely by enlightened guesswork. Some relevant considerations are as follows.

(i) As in Wiens (1986), a general principle appears to be that for sufficiently small ϵ, ψ_0 should differ from $\tilde{\xi}$ only near the local extrema of $J(\tilde{\xi})$, and that here we should have $J(\psi_0) =$ constant, with the constant being less extreme than that attained by $J(\tilde{\xi})$. In line with condition (2) of Lemma 2 we should have $\tilde{f}_0 > \tilde{g}, \tilde{F}_0 - \tilde{G}$ increasing from -2ϵ to 2ϵ, near the local minima of $J(\tilde{\xi})$; and $\tilde{f}_0 < \tilde{g}, \tilde{F}_0 - \tilde{G}$ decreasing from 2ϵ to -2ϵ, near the local maxima. As ϵ increases, the regions of constancy of $J(\psi_0)$ coalesce.

(ii)Under conditions analogous to (G1)–(G4), solutions for symmetric $\tilde{G}$ were obtained by Wiens (1986). In particular, $J(\tilde{\xi})$ was assumed symmetric and unimodal. The solutions obtained there have two forms — form 1, valid for "small" ϵ, and form 2, for "large" ϵ. In the present case, suppose that the two halves of $J(\tilde{\xi})$ — to the right and left of M — are considered separately, and each is extended to all of $\mathbb{R}$ by symmetry around M. Each then generates a solution which is symmetric around M, and it seems plausible that we should be able to piece together these two solutions. This does indeed turn out to be the case, although the piecing together is not necessarily done at M. For small values of ϵ we piece together two form-1 solutions, for medium ϵ a form-1 with a form-2, and for large ϵ we piece together two form-2 solutions.

It is of course still necessary to verify the conditions of Theorem 4. Some of these verifications are presented below; for the remainder the reader is referred to Wu (1990). In each case, χ_0 and f_0 may be recovered from ψ_0 and $\tilde{f}_0$ through (2.4) and (2.5).

CASE 1 (Small ϵ). *There exists $\epsilon_0(G)$ such that for $0 < \epsilon \le \epsilon_0(G)$, the Fisher information for scale is minimized by that F_0 with*

$$\psi_0(y) = \begin{cases} \tilde{\xi}(b_1), & -\infty < y \le b_1, \\ \tilde{\xi}(y), & b_1 \le y \le a_2, \\ \delta \tan\left(\dfrac{\delta}{2}(y-w)\right), & a_2 \le y \le b_2, \\ \tilde{\xi}(y), & b_2 \le y \le a_3, \\ \tilde{\xi}(a_3), & a_3 \le y < \infty, \end{cases}$$

$$\tilde{f}_0(y) = \begin{cases} \tilde{g}(b_1)e^{\psi_0(b_1)(b_1-y)}, & -\infty < y \le b_1, \\ \tilde{g}(y), & b_1 \le y \le a_2, \\ \dfrac{\tilde{g}(a_2)\cos^2\{(\delta/2)(y-w)\}}{\cos^2\{(\delta/2)(a_2-w)\}}, & a_2 \le y \le b_2, \\ \tilde{g}(y), & b_2 \le y \le a_3, \\ \tilde{g}(a_3)e^{\psi_0(a_3)(y-a_3)}, & a_3 \le y < \infty, \end{cases}$$

The constants $b_1 < a_2 < b_2 < a_3, \delta$ and w are determined by

(a) continuity of ψ_0 at a_2 and b_2, and of $\tilde{f}_0$ at b_2,
(b) $\tilde{F}_0(b_1) = \tilde{G}(b_1) + 2\epsilon, \tilde{F}_0(b_2) = \tilde{G}(b_2) - 2\epsilon, \tilde{F}_0(\infty) = 1$,

and satisfy as well

(c) $a_2 < M < b_2$,
(d) $\delta^2 \ge \max\left(J(\tilde{\xi})(a_2), J(\tilde{\xi})(b_2)\right)$.

Then $B_U = [b_1, a_2], B_L = [b_2, a_3]$.

Note that if (a) is satisfied, then $\tilde{f}_0$ and ψ_0 are (absolutely) continuous, so that (S2) of Theorem 4 is satisfied. Condition (S3)(iv) follows from (c) above. In the notation of

(S3) we have

$$n = 3, \qquad \lambda_1 = -\tilde{\xi}^2(b_1), \qquad \lambda_2 = \delta^2, \qquad \lambda_3 = -\tilde{\xi}^2(a_3).$$

The requirements

$$\lambda_3 \le J(\tilde{\xi})(a_3), \qquad \lambda_1 \le J(\tilde{\xi})(b_1)$$

of (S3)(iii) are then immediate consequences of Lemma 3. To satisfy (S1) and the remaining parts of (S3)(iii), it suffices if (d) holds, and if

$$\tilde{f}_0(y) \ge \tilde{g}(y), \qquad y \in (-\infty,\, b_1) \cup (a_3,\, \infty); \tag{3.1}$$

$$\tilde{f}_0(y) \le \tilde{g}(y), \qquad y \in (a_2,\, b_2). \tag{3.2}$$

The requirement (3.1) follows from the equality of $\tilde{f}_0$ and $\tilde{g}$ at b_1 and a_3, together with the inequalities

$$\psi_0(y) \ge \tilde{\xi}(y), \quad y < b_1; \qquad \psi_0(y) \le \tilde{\xi}(y), \quad y > a_3,$$

and the identity at the end of Remark 5 above.

For (3.2) note that if (d) holds, then ψ_0' exceeds $\tilde{\xi}'$ at a_2 and b_2, at which point ψ_0 equals $\tilde{\xi}$. There is then a point $M_0 \in (a_2, b_2)$ at which $\psi_0(M_0) = \tilde{\xi}(M_0)$. Conditions (c), (d) together imply that $\psi_0 - \tilde{\xi}$ is decreasing at any zero in (a_2, b_2) — see Wu (1990) for details. The point M_0 is thus unique, with $\psi_0 > \tilde{\xi}$ on (a_2, M_0), $\psi_0 < \tilde{\xi}$ on (M_0, b_2). Now (3.2) follows in the same manner as (3.1).

For the verification that there exist constants satisfying (a)–(d) above, see Wu (1990).

For "medium ϵ" there are two possibilities, according to whether the form-1 solution is placed to the right or to the left of the form-2 solution.

CASE 2 (Medium ϵ). *For a range $\epsilon_0(G) \le \epsilon \le \epsilon_1(G)$, the Fisher information for scale is minimized either by that $\tilde{f}_0(y)$ with density $\tilde{f}_0(y)$ and score function $\psi_0(y)$ as given below, or by $1 - \tilde{F}_0(-y)$, with density $f_0(-y)$ and score function $-\psi_0(-y)$. If one form minimizes the information in a neighbourhood of $\tilde{G}(y)$, then the other minimizes the information in a neighbourhood of $1 - \tilde{G}(-y)$:*

$$\psi_0(y) = \begin{cases} \tilde{\xi}(b_1), & -\infty < y \le b_1, \\ \tilde{\xi}(y), & b_1 \le y \le a_2, \\ \delta \tan\left(\dfrac{\delta}{2}(y - w)\right), & a_2 \le y \le a_3, \\ \delta \tan\left(\dfrac{\delta}{2}(a_3 - w)\right), & a_3 \le y \le \infty, \end{cases} \tag{3.3}$$

$$\tilde{f}_0(y) = \begin{cases} \tilde{g}(b_1) e^{\psi_0(b_1)(b_1 - y)}, & -\infty < y \le b_1, \\ \tilde{g}(y), & b_1 \le y \le a_2, \\ \dfrac{\tilde{g}(a_2) \cos^2\{(\delta/2)(y - w)\}}{\cos^2\{(\delta/2)(a_2 - w)\}}, & a_2 \le y \le a_3, \\ \tilde{g}(a_3) e^{-\psi_0(a_3)(y - a_3)}, & a_3 \le y < \infty. \end{cases} \tag{3.4}$$

The constants $b_1 < a_2 < a_3, \delta$, and w are determined by

(a) *continuity of ψ_0 at a_2 and $\tilde{f}_0$ at a_3,*

(b) $\tilde{F}_0(b_1) = \tilde{G}(b_1) + 2\epsilon$, $\tilde{F}_0(a_3) = \tilde{G}(a_3) - 2\epsilon$, $\tilde{F}_0(\infty) = 1$,

and satisfy as well

(c) $a_2 < M$, $\psi_0(a_3) \leq \tilde{\xi}(a_3)$,
(d) $\delta^2 \geq J(\tilde{\xi})(a_2)$.

Then $B_U = [b_1, a_2], B_L = \{a_3\}$.

Given the existence of constants satisfying (a)–(d), the verification of the conditions of Theorem 4 is very similar to that in Case 1. Note that in Case 1, it was required by (S3)(iv) that M be less than b_2. In Case 2, $b_2 = a_3$, and this requirement is replaced by (S4)(ii) in condition (c) above.

CASE 3 (Large ϵ). *For* $\epsilon_1(G) \leq \epsilon \leq 0.25$, *the Fisher information for scale is minimized by that* $\tilde{F}_0$ *with*

$$\psi_0(y) = \begin{cases} \delta \tan\left(\frac{\delta}{2}(b_1 - w)\right), & -\infty < y \leq b_1, \\ \delta \tan\left(\frac{\delta}{2}(y - w)\right), & b_1 \leq y \leq a_3, \\ \delta \tan\left(\frac{\delta}{2}(a_3 - w)\right), & a_3 \leq y < \infty, \end{cases}$$

$$\tilde{f}_0(y) = \begin{cases} \tilde{g}(b_1)e^{\psi_0(b_1)(b_1 - y)}, & -\infty < y \leq b_i, \\ \dfrac{\tilde{g}(a_2)\cos^2\{(\delta/2)(y - w)\}}{\cos^2\{(\delta/2)(b_1 - w)\}}, & b_1 \leq y \leq a_3, \\ \tilde{g}(a_3)e^{-\psi_0(a_3)(y - a_3)}, & a_3 \leq y < \infty. \end{cases}$$

The constants $b_1 < a_3, \delta$, *and* w *are determined by*

(a) continuity of $\tilde{f}_0$ *at* a_3,
(b) $\tilde{F}_0(b_1) = \tilde{G}(b_1) + 2\epsilon$, $\tilde{F}_0(a_3) = \tilde{G}(a_3) - 2\epsilon$, $\tilde{F}_0(\infty) = 1$,

and satisfy as well

(c) $\psi_0(b_1) \geq \tilde{\xi}(b_1)$, $\psi_0(a_3) \leq \tilde{\xi}(a_3)$.

Then $B_U = \{b_1\}, B_L = \{a_3\}$.

EXAMPLE 1. In the following three cases, the preceding theory applies and the "medium-ϵ" solution is as in (3.3), (3.4):

$$G_1 = \Phi, \qquad g_1(x) = (2\pi)^{-1/2}e^{-x^2/2}, \qquad \tilde{\xi}_1(y) = e^{2y} - 1;$$

$$G_2 = \text{Logistic}, \quad g_2(x) = \tfrac{1}{4}\,\text{sech}^2(x/2), \qquad \tilde{\xi}_2(y) = e^y \tanh(e^y/2) - 1;$$

$$G_3 = \text{Laplace}, \quad g_3(x) = \tfrac{1}{2}\,e^{-|x|}, \qquad \tilde{\xi}_3(y) = e^y - 1.$$

We find that $\epsilon_0(\Phi) = 0.00205, \epsilon_1(\Phi) = 0.0267$. See Table 1 for other numerical values if $G = \Phi$, and Figures 1–3 for graphs of $\chi_0(x)$ in this case, For G_3, see Table 2.

Note that for very small values of ϵ, χ_0 differs significantly from ξ only in the tails. For larger ϵ, it evidently becomes more important to guard against inliers in the data.

EXAMPLE 2. An Associate Editor has pointed out that if $G(x)$ is the Cauchy distribution, then $\tilde{G}(y)$ is the hyperbolic-secant distribution, which is itself symmetric, with $J(\tilde{\xi})$ a

TABLE 1: Numerical values for $K_{2\epsilon}(\tilde{G})$, $G = \Phi$.

ϵ	b_1	a_2	b_2	a_3	δ	ω	$I(F_0;1) = I_*(\tilde{F}_0)$
0.0001	−2.63109	0.27763	0.76640	1.18743	2.7100	0.08103	1.9710
0.0005	−2.09720	0.15529	0.83898	1.06022	2.6100	0.04897	1.9050
0.0008	−1.94212	0.10796	0.86416	1.01523	2.5680	0.03505	1.8670
0.0010	−1.86821	0.08618	0.87505	0.99399	2.5490	0.02840	1.8450
0.0018	−1.67611	0.00896	0.91188	0.93452	2.4780	0.00315	1.7690
0.00205	−1.63373	0.00894	0.92068	0.92068	2.4610	−0.00309	1.7480
0.004	−1.41717	−0.18252		0.87383	2.3410	−0.04605	1.6090
0.010	−1.12701	−0.38287		0.82154	2.0710	−0.13868	1.3180
0.020	−0.91554	−0.67276		0.78709	1.7870	−0.23358	1.0090
0.025	−0.84957	−0.79208		0.77932	1.6830	−0.26750	0.8947
0.0267	−0.83034	−0.83034		0.77749	1.6510	−0.27777	0.8600
0.030	−0.87731			0.77473	1.5930	−0.29642	0.7927
0.040	−1.00923			0.77288	1.4410	−0.34559	0.6387
0.100	−1.64351			0.83508	0.8567	−0.56169	0.1721
0.200	−3.08347			1.05883	0.2847	−1.07762	0.0067
0.230	−4.18646			1.18723	0.1384	−1.52746	0.0007

TABLE 2: Numerical values for $K_{2\epsilon}(\tilde{G}_3)$ and $K_{2\epsilon}(\tilde{G}_4)$; G_3 = Laplace, G_4 = Negative Exponential.

ϵ	b_1	a_2	b_2	a_3	δ	ω	$I_*(\tilde{F}_0)$
0.00005	−4.26158	0.33002	1.00393	1.99606	1.6900	0.06154	0.9940
0.00050	−3.12357	0.09531	1.16814	1.69028	1.6290	0.01842	0.9631
0.00150	−2.58495	−0.09310	1.27256	1.51381	1.5770	−0.02169	0.9227
0.00250	−2.33718	−0.19760	1.32840	1.42335	1.5450	−0.04790	0.8896
0.00300	−2.24999	−0.23889	1.34963	1.38954	1.5320	−0.05875	0.8746
0.00325	−2.21092	−0.25812	1.35918	1.37447	1.5250	−0.06426	0.8662
0.00400	−2.11196	−0.31047		1.34859	1.5090	−0.07833	0.8323
0.00500	−2.00619	−0.37571		1.32522	1.4860	−0.09623	0.7983
0.01000	−1.68201	−0.63885		1.25390	1.3890	−0.16746	0.7301
0.02500	−1.27047	−1.16603		1.17927	1.1910	−0.28598	0.5131
0.02730	−1.21200	−1.23100		1.17465	1.1680	−0.29726	0.4893
0.05000	−1.50824			1.16938	0.9742	−0.38390	0.3092
0.10000	−2.01741			1.26271	0.6648	−0.52768	0.1090
0.15000	−2.58893			1.41658	0.4315	−0.69293	0.0309
0.20000	−3.46734			1.63803	0.2341	−0.97480	0.0016
0.22500	−4.29769			1.81271	0.1374	−1.27656	0.0008
0.24500	−6.16582			2.11251	0.0456	−2.03070	0.0002

strictly decreasing function of $|y|$. The theory of Wiens (1986) thus applies directly to this case. The least favourable $\tilde{F}_0$ is symmetric, and $\epsilon_0(G) = \epsilon_1(G)$ — there is no "medium-ϵ" form to the solution.

EXAMPLE 3. If $G(x) = G_4(x) = 1 - e^{-x}, x > 0$, then the distribution $\tilde{G}_4(y)$ of $\log X$ is the same as $\tilde{G}_3(y)$ of Example 1. Solutions $\tilde{F}_0(y)$ in neighbourhoods $\mathcal{K}_{2\epsilon}(\tilde{G}_4)$ are then the same as in neighbourhoods $\mathcal{K}_{2\epsilon}(\tilde{G}_3)$, detailed in Table 2. In the original units, the least favourable density in $\mathcal{K}_{2\epsilon}(G_4)$ is recovered from

$$f_0(x) = x^{-1}\tilde{f}_0(\log x), \qquad x > 0,$$

rather than through (2.4).

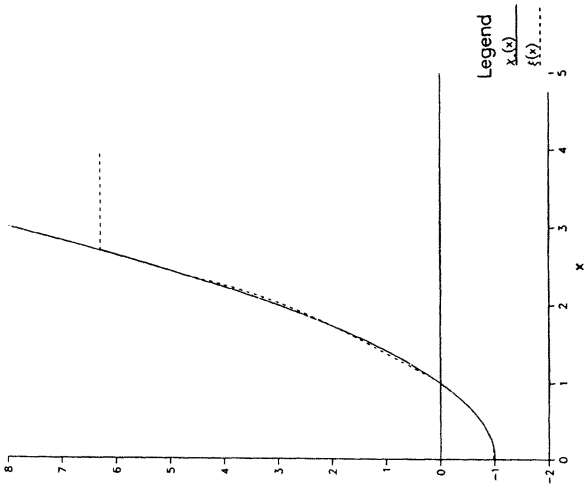

FIGURE 1: Graph of $\chi_0(x)$; $G = \Phi$, small $\epsilon(\epsilon = 0.001)$.

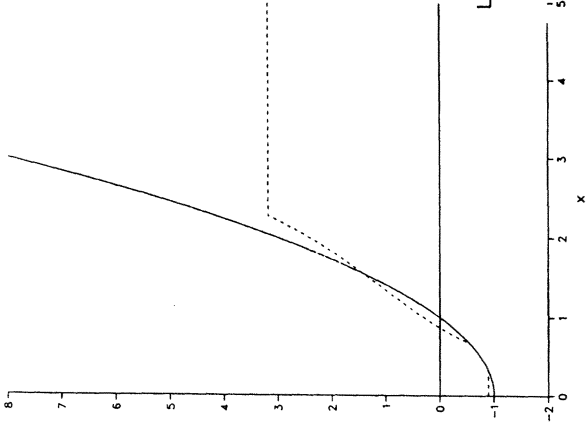

FIGURE 2: Graph of $\chi_0(x)$; $G = \Phi$, medium $\epsilon(\epsilon = 0.1)$.

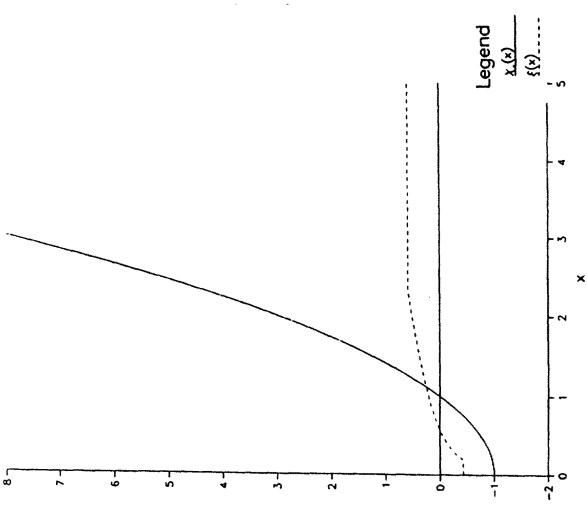

FIGURE 3: Graph of $\chi_0(x)$; $G = \Phi$, large $\epsilon(\epsilon = 0.1)$.

Thall (1979) presents a different solution to the problem of minimizing information for scale in $\mathcal{K}_\epsilon(G_4)$. He presents a solution similar to $1 - \tilde{F}_0(-y)$ in our Case 2, and claims that it is valid for all $\epsilon \leq 0.0095$. Unfortunately, his solution fails to satisfy (S4)(ii). The $\tilde{F}_0$ constructed by Thall has $\tilde{f}_0 < \tilde{g}$ at the single point at which $\tilde{F}_0 = \tilde{G}_4 + \epsilon$. Thus, this $\tilde{F}_0$ exceeds $\tilde{G}_4 + \epsilon$ to the left of this point, and so fails to belong to $\mathcal{K}_\epsilon(\tilde{G}_4)$.

ACKNOWLEDGEMENT

The authors would like to thank an Associate Editor, whose insightful comments improved the results and their presentation.

REFERENCES

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.

Collins, J., and Wiens, D. (1989). Minimax properties of M-, R- and L-estimators of location in Lévy neighbourhoods. *Ann. Statist.*, 17, 327–336.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, 35, 73–101.

Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

Sacks, J., and Ylvisaker, D. (1972). A note on Huber's robust estimation of a location parameter. *Ann. Math. Statist.*, 43, 1068–1075.

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

Thall, P.R. (1979). Huber-sense robust M-estimation of a scale parameter, with application to the exponential distribution. *J. Amer. Statist. Assoc.*, 74, 147–152.

Vandelinde, V.D. (1979). Robust techniques in communication. *Robustness in Statistics* (R. Launer and G. Wilkinson, *eds.*), Academic Press, New York, 177–200.

Wiens, D.P. (1986). Minimax variance M-estimators of location in Kolmogorov neighbourhoods. *Ann. Statist.*, 14, 724–732.

Wiens, D.P., and Wu, K.H.E. (1990). Asymptotic minimax properties of M-estimators of scale. *Statist. Probab. Lett.*, 10, 363–368.

Wu, K.H.E. (1990). Asymptotic minimax properties of M-estimators for scale. Ph.D. Thesis, University of Alberta.

Received 10 October 1989
Revised 4 July 1990
Accepted 13 September 1990

Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta T6G 2G1
Department of Statistics
The Chinese University of Hong Kong
Shatin, N.T.
Hong Kong