

## MINIMAX VARIANCE M-ESTIMATORS IN $\varepsilon$ -CONTAMINATION MODELS<sup>1</sup>

BY JOHN R. COLLINS AND DOUGLAS P. WIENS

*University of Calgary and Dalhousie University*

In the framework of Huber's theory of robust estimation of a location parameter, minimax variance M-estimators are studied for error distributions with densities of the form  $f(x) = (1 - \varepsilon)h(x) + \varepsilon g(x)$ , where  $g$  is unknown. A well-known result of Huber (1964) is that when  $h$  is strongly unimodal, the least informative density  $f_0 = (1 - \varepsilon)h + \varepsilon g_0$  has exponential tails. We study the minimax variance solutions when the known density  $h$  is not necessarily strongly unimodal, and definitive results are obtained under mild regularity conditions on  $h$ . Examples are given where the support of the least informative contaminating density  $g_0$  is a set of form: (i)  $(-b, -a) \cup (a, b)$  for some  $0 < a < b < \infty$ ; (ii)  $(-a, a)$  for some  $0 < a < \infty$ ; and (iii) a countable collection of disjoint sets. Minimax variance problems for multivariate location and scale parameters are also studied, with examples given of least informative distributions that are substochastic.

**1. Introduction and summary.** In Huber's (1964) theory of robust estimation of a location parameter, the minimax variance M-estimator has score function  $\psi_0 = -f'_0/f_0$  corresponding to the least informative density  $f_0$  in a convex class  $\mathcal{F}$ . An important model for the class  $\mathcal{F}$  of unknown error distributions is the  $\varepsilon$ -contamination model: let  $\varepsilon$ ,  $0 < \varepsilon < 1$ , be known and let  $h$  be a fixed known density function symmetric about 0. Then the model is that  $\mathcal{F}$  is the class of densities of form

$$(1.1) \quad f(x) = (1 - \varepsilon)h(x) + \varepsilon g(x)$$

where  $g$  is an unknown density symmetric about 0.

It is a well-known result of Huber (1964) that when the density  $h$  in the  $\varepsilon$ -contamination model is strongly unimodal, then the least favorable  $f_0$  has exponential tails and the corresponding  $\psi_0 = -f'_0/f_0$  is given by

$$(1.2) \quad \begin{aligned} \psi_0(x) &= \min\{-h'(x)/h(x), k\} & \text{for } x \geq 0, \\ &= -\psi_0(-x) & \text{for } x < 0, \end{aligned}$$

where  $k$  depends on  $\varepsilon$ . In this paper we study the form of the minimax solution for general  $\varepsilon$ -contamination models, i.e., for cases where  $h$  is not necessarily strongly unimodal. For example, when  $h$  is the Cauchy density  $h(x) = [\pi(1 + x^2)]^{-1}$ , then the least informative  $f_0$  cannot possibly have exponential tails

---

Received December 1983; revised October 1984.

<sup>1</sup> Research supported by Natural Sciences and Engineering Research Council of Canada Grants A-4499 and A-8603.

AMS 1980 subject classifications. Primary 62F35; secondary 62F12.

Key words and phrases. Robust estimation, M-estimators, minimax variance.

because the tails of the uncontaminated Cauchy  $h$  are heavier than exponential. The investigation of general  $\varepsilon$ -contamination models was motivated by a conjecture of Joiner and Hall (1980). The results reported here form part of the Ph.D. dissertation of Wiens (1982).

Huber's necessary and sufficient condition for  $f_0$  to be the least informative density in  $\mathcal{F}$  is that

$$(1.3) \quad \int [-2\psi_0(f' - f'_0) - \psi_0^2(f - f_0)] dx \geq 0$$

for all absolutely continuous  $f$  in  $\mathcal{F}$ . Writing  $f = (1 - \varepsilon)h + \varepsilon g$  and  $f_0 = (1 - \varepsilon)h + \varepsilon g_0$ , an integration by parts (provided that  $\psi_0$  is suitably regular) yields  $\int [2\psi'_0 - \psi_0^2](g - g_0) dx \geq 0$  for all  $g$ . From this it follows that the support of  $g_0$  must be the set of values of  $x$  on which  $2\psi'_0(x) - \psi_0^2(x)$  attains its minimum value. Further necessary and sufficient conditions can be deduced which allow one to find the minimax solution under very mild regularity conditions on  $h$ .

The general results are given in Section 2. Examples are given in Section 3, including the case where  $h$  is Cauchy and some other somewhat surprising examples. In Section 4, the theory is extended to the case of redescending M-estimators.

**2. The general theory.** We present an asymptotic minimax variance theory general enough to include as special cases both the one-dimensional and  $m$ -dimensional location and scale (or scatter) models with contaminating distributions restricted to be symmetric. The theory is not general enough to include other closely related asymptotic optimality criteria such as the "change-of-variance curve" approach of Hampel, Rousseeuw and Ronchetti (1981) (which also yields solutions for which  $2\psi'_0 - \psi_0^2$  is constant on some intervals). Nor does the theory include some important minimax problems arising in robust regression (see, e.g., Bickel, 1984; and Huber, 1983).

Throughout this section, let  $\nu$  and  $\eta$  denote fixed functions satisfying the following conditions:

- (C.1)  $\nu: [0, \infty) \rightarrow [0, \infty)$  and  $\eta: [0, \infty) \rightarrow [0, \infty)$  are absolutely continuous functions, which are positive and continuously differentiable on  $(0, \infty)$ ;
- (C.2)  $|\nu'(x)/\nu(x)|$  is bounded as  $x \rightarrow \infty$ ; and
- (C.3)  $\eta(x)$  is bounded away from 0 as  $x \rightarrow \infty$ .

Let the function  $\sigma: [0, \infty) \rightarrow [0, \infty)$  be defined by  $\sigma = \eta\nu$ . Let  $\mathcal{G}$  be the class of functions defined by

$$\mathcal{G} = \left\{ G \left| \int_0^x \nu(y) dG(y) \text{ is a (possibly substochastic)} \right. \right. \\ \left. \left. \text{distribution function on } \mathbb{R}^+ \text{ with } \int_0^\infty \nu(y) dG(y) \leq \frac{1}{2} \right\}.$$

For each  $G$  in  $\mathcal{G}$ , define  $G_\nu$  on  $[0, \infty)$  by

$$G_\nu(x) = \int_0^x \nu(y) dG(y),$$

and define  $\mathcal{G}_\nu$  by  $\mathcal{G}_\nu = \{G_\nu \mid G \in \mathcal{G}\}$ . Also denote by  $\mathcal{G}'$  the subclass of  $G$ 's in  $\mathcal{G}$  which have an absolutely continuous derivative  $g$ , and let  $\mathcal{G}'_\nu = \{G_\nu \mid G \in \mathcal{G}'\}$ .

Let  $\epsilon, 0 < \epsilon < 1$ , be fixed, and let  $H$  be a fixed member of  $\mathcal{G}'$ , having additional properties to be specified later. Define  $\mathcal{F}$  by

$$(2.1) \quad \mathcal{F} = \{F \mid F = (1 - \epsilon)H + \epsilon G \text{ for some } G \in \mathcal{G}\}.$$

Also define  $\mathcal{F}_\nu$  by  $\mathcal{F}_\nu = \{F_\nu \mid F \in \mathcal{F}\}$  (where  $F_\nu(x) = \int_0^x \nu(y) dF(y)$ ) and define  $\mathcal{F}' = \mathcal{F} \cap \mathcal{G}'$  and  $\mathcal{F}'_\nu = \mathcal{F}' \cap \mathcal{G}'_\nu$ .

Let  $\Psi$  denote the class of functions  $\psi: [0, \infty) \rightarrow \mathbb{R}$  which are continuous and have a piecewise continuous derivative. For each  $\psi \in \Psi$  and  $F \in \mathcal{F}'$ , with absolutely continuous density  $f$ , define the functional  $V(\psi, f)$  by

$$(2.2) \quad V(\psi, f) = \left( \int_0^\infty \psi^2(x) \eta(x) f(x) \nu(x) dx \right) / \left( 2 \left[ \int_0^\infty \psi(x) \eta(x) f'(x) \nu(x) dx \right]^2 \right).$$

Then the general minimax problem is to find a  $\psi$  in  $\Psi$  which minimizes  $\sup\{V(\psi, f) : F \in \mathcal{F}'\}$ .

**EXAMPLE 2.1.** Huber (1964) considers the estimation of  $\theta \in \mathbb{R}$  when  $X_1, \dots, X_n$  is a random sample from an (approximately specified) distribution  $F(\cdot - \theta)$  using M-estimators, i.e., solutions of  $\sum_{i=1}^n \psi(X_i - \theta) = 0$ . One way to force the M-estimator to be a consistent estimator of  $\theta$  is to impose the side conditions that the distribution  $F$  be symmetric about 0 and that the function  $\psi$  be skew-symmetric about 0. Then, under regularity conditions, Huber (1964) shows that the (consistent) M-estimator of  $\theta$  is asymptotically normal with asymptotic variance

$$\left( \int_{-\infty}^\infty \psi^2(x) f(x) dx \right) / \left( \left[ \int_{-\infty}^\infty \psi(x) f'(x) dx \right]^2 \right)$$

when  $F$  has absolutely continuous density  $f$ . By symmetry, the variance functional can be written as  $V(\psi, f)$  (formula (2.2)) when  $\eta(x) = \nu(x) \equiv 1$ . With this choice of  $\eta$  and  $\nu$  (which trivially satisfy (C.1), (C.2) and (C.3)), the general minimax problem reduces to finding the  $\psi$  which minimizes  $\sup\{\int \psi^2 f / (\int \psi f')^2\}$  as  $f$  varies over all absolutely continuous symmetric densities  $f$  of the form  $f = (1 - \epsilon)h + \epsilon g$ , where  $h = H'$  is a fixed density symmetric about 0. The solution was obtained by Huber (1964) for all cases in which  $h$  is twice differentiable and  $-\log h(x)$  is convex on the support of  $H$ . (Note that if  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  is a skew-symmetric function which has a piecewise continuous derivative and which is continuous except

perhaps at  $x = 0$ , then the restriction of  $\psi$  to  $[0, \infty)$  is in the class  $\Psi$  defined above.)

EXAMPLE 2.2. Huber (1977) and Maronna (1976) considered the following problem (see also Chapter 8 of Huber, 1981). Let the random vector  $\mathbf{Y} \in \mathbb{R}^m$  have a spherically symmetric density  $g(\mathbf{y}) = g(|\mathbf{y}|)$ . Let

$$(2.3) \quad \nu(x) = \frac{1}{2}mC_m|x|^{m-1},$$

where  $C_m$  denotes the volume of the unit sphere in  $\mathbb{R}^m$ . Then with  $G(x) = \int_0^x g(y) dy$  and  $G_\nu(x) = \int_0^x \nu(y) dG(y)$  for  $x \geq 0$ , one can see that the random variable  $X = |\mathbf{Y}|$  has distribution function  $2G_\nu$ . Now assume that a nondegenerate affine transformation  $\mathbf{y} \rightarrow V\mathbf{y} + \mathbf{t}$  has been applied, and that the problem is to estimate the unknown location vector  $\mathbf{t} \in \mathbb{R}^m$  and the unknown scatter (scale) matrix  $V$ . Huber (1977) and Maronna (1976) proposed estimating  $\mathbf{t}$  and  $V$  by solving a system of equations which are shown to result in consistent asymptotically normal estimators. Furthermore, the location and scatter coordinates of the limiting covariance matrix are seen to be asymptotically independent and are determined up to scalar-valued (because of spherical symmetry) functions of  $u_1$  and  $f$  (for location) and  $u_2$  and  $f$  (for scale), where  $u_1$  and  $u_2$  are "influence functions" for location and scale, respectively. In particular, if  $\nu$  is given by (2.3) and  $\eta_i$  is defined by

$$\eta_i(x) = (2/(m + 2))^i(x^{2i}/m) \quad \text{for } i = 0, 1,$$

then the minimax problem (2.2) can be seen (up to some differences in parameterization) to be equivalent to the minimax variance problems of Huber (1977) with  $i = 0$  and  $i = 1$  corresponding to location and scale, respectively. (Note that  $\nu$ ,  $\eta_0$  and  $\eta_1$  clearly satisfy conditions (C.1), (C.2) and (C.3).)

We remark that the restriction to spherically symmetric contaminating distributions, which reduces a multivariate problem to a univariate one, is quite severe. But the restriction is necessary to make the minimax variance criterion meaningful (by eliminating bias). For a modification of the theory which allows contamination which is symmetric in a central region but asymmetric further out, see Section 4. For a different approach to the robust estimation problem that places no restrictions on the distribution of outliers, see Donoho and Huber (1983).

Returning to the general problem of finding the  $\psi$  which minimizes  $\sup\{V(\psi, f): F \in \mathcal{F}'\}$  with  $V(\psi, f)$  given by (2.2), we give a modification of Huber's definition of Fisher information which is appropriate to our problem. For fixed  $\nu$ ,  $\eta$  and  $\sigma = \nu\eta$  satisfying (C.1), (C.2) and (C.3), and for all  $F_\nu \in \mathcal{F}_\nu$ , define the functional  $I(\cdot)$  by

$$I(F_\nu) = \sup_\psi \left[ 2 \left( \int_0^\infty (\psi\sigma)' dF \right)^2 / \int_0^\infty \psi^2 \sigma dF \right],$$

where the sup is over the set  $C_k^1$  of all continuously differentiable functions with compact support satisfying  $\int_0^\infty \psi^2 \sigma dF > 0$ .

Simple modifications to the proofs of Theorem 4.2, and Propositions 4.3, 4.5 of Huber (1981) give:

**THEOREM 1.**

- (A) *The following are equivalent:*
  - (i)  $I(F_\nu) < \infty$ .
  - (ii)  $F$  has an absolutely continuous derivative  $f$  satisfying  $\int_0^\infty (f'/f)^2 f \sigma dx < \infty$ .

*In either case,  $I(F_\nu) = 2 \int_0^\infty (f'/f)^2 f \sigma dx$ .*
- (B) *There is an  $F_\nu^0 \in \mathcal{F}_\nu$  minimizing  $I(F_\nu)$ . Define  $F_0$  by  $F_\nu^0(x) = \int_0^x \nu(y) dF_0(y)$ .*
- (C) *If  $0 < I(F_\nu^0) < \infty$ , and the set where  $f_0 = F_0'$  is strictly positive is convex, then  $F_\nu^0$  is unique.*

Motivated by Theorem 1, we impose some further conditions on the function  $H$  in our model (2.1):

- (H.1)  $0 < I(H_\nu) < \infty$ ;
- (H.2)  $h(=H')$  is strictly positive on  $[0, \infty)$ ;
- (H.3)  $\lim_{x \rightarrow \infty} (h\nu)(x) = 0$ ;
- (H.4) the function  $\zeta$ , defined by  $\zeta = -h'/h$  is absolutely continuous and continuously differentiable; and
- (H.5)  $\lim_{x \downarrow 0} (\zeta\sigma)(x) \geq 0$ .

The conditions imposed on  $H$  ensure that there is a unique  $F_\nu^0 \in \mathcal{F}_\nu'$  minimizing  $I(F_\nu)$  over  $\mathcal{F}_\nu$ . Since  $I(F_\nu)$  is a convex functional of  $F_\nu$  (see Lemma 4.4, Huber, 1981),  $F_\nu^0$  minimizes  $I(F_\nu)$

iff  $(d/dt)I(F_\nu^t)|_{t=0} \geq 0$  for all  $F_\nu^t = (1 - t)F_\nu^0 + tF_\nu^1$  with  $F_\nu^1 \in \mathcal{F}_\nu'$ .

Performing the differentiation and setting  $\psi_0 = -f_0'/f_0$  gives the necessary and sufficient condition

$$(2.4) \quad 0 \leq \int_0^\infty [2(f_0' - f')\psi_0 + (f_0 - f)\psi_0^2]\sigma dx$$

for all  $f = F'$  with  $F_\nu \in \mathcal{F}_\nu'$ .

We can now relate the information and variance functionals via Theorem 2 of Huber (1964).

**THEOREM 2.** *Under the assumptions (H.1) and (H.2), there is a unique  $F_\nu^0 \in \mathcal{F}_\nu'$  minimizing  $I(F_\nu)$ .*

- (i) *If  $\psi_0 = -f_0'/f_0$  is contained in  $\Psi$ , then  $(\psi_0, f_0)$  is a saddlepoint of  $V(\psi, f)$  in  $\Psi \times \mathcal{F}_\nu'$ :*

$$(2.5) \quad V(\psi_0, f) \leq 1/I(F_\nu^0) = V(\psi_0, f_0) \leq V(\psi, f_0),$$

for all  $\psi \in \Psi$  and  $F_\nu \in \mathcal{F}_\nu'$ .

- (ii) Conversely, if  $(\psi_0, f_0)$  is a saddlepoint, and  $\Psi$  contains a nonzero multiple of  $-f'_0/f_0$ , then  $I(F^0_\nu) \leq I(F_\nu)$  for all  $F_\nu \in \mathcal{F}_\nu$ ,  $F^0_\nu$  is uniquely determined, and  $\psi_0$  is  $[F^0_\nu]$ -equivalent to a multiple of  $-f'_0/f_0$ .
- (iii) Necessary and sufficient for  $F^0_\nu$  to minimize  $I(F_\nu)$  is that (2.4) above be satisfied.

Note that by writing  $f_0 = (1 - \varepsilon)h + \varepsilon g_0$ , the necessary and sufficient condition (2.4) can be rewritten as

$$(2.6) \quad 0 \leq \int_0^\infty \left\{ 2[(g_0\nu)' - (g\nu)']\psi_0\eta + (g_0\nu - g\nu) \left[ -2\psi_0\eta \frac{\nu'}{\nu} + \psi_0^2\eta \right] \right\} dx$$

for all  $G \in \mathcal{G}'$ . Our main theorem (see Theorem 3) gives a further set of necessary and sufficient conditions which make possible the explicit determination of the minimax  $\psi_0$  corresponding to a given  $H$  satisfying (H.1)–(H.5). For convenient use in the statement and proof of the theorem, we define the functional  $J$ , for all differentiable  $\psi: [0, \infty) \rightarrow \mathbb{R}$ , by:

$$J(\psi) = (2\psi' - \psi^2 + 2\psi(\sigma'/\sigma))\eta.$$

It turns out that the minimax  $\psi_0$  agrees with  $\zeta = -h'/h$  except on a (finite or countable) collection of disjoint open intervals  $\{B_{\lambda,j}\}$  where  $\psi_0$  agrees with a solution  $\xi$  of  $J(\xi) \equiv -\lambda$ . Since on each  $B_{\lambda,j}$  the solution typically has the same parametric form but with parameter values depending on  $B_{\lambda,j}$ , we introduce an index  $\{\omega_j\}$  to keep track of the version of the solution on  $B_{\lambda,j}$ , and we write  $J(\xi(x; \omega_j, \lambda)) \equiv -\lambda$  for  $x \in B_{\lambda,j}$ .

**THEOREM 3.** Under assumptions (C.1) through (C.3), and (H.1)–(H.5), there is a unique  $F^0_\nu \in \mathcal{F}'_\nu$  minimizing  $I(F_\nu)$  over  $\mathcal{F}_\nu$ .

(A) In order that  $F^0_\nu$  minimize  $I(F_\nu)$ , the following are necessary and sufficient:

(P.1) The function  $\psi_0\eta = -(f'_0/f_0)\eta$  is bounded, absolutely continuous, and piecewise continuously differentiable on  $(0, \infty)$ .

(P.2)  $f_0\nu(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

There exists  $\lambda \geq 0$ , and a set  $B_\lambda = \cup_{j=1}^{N(\lambda)} B_{\lambda,j} \subseteq (0, \infty)$ , where  $N(\lambda) \leq \infty$  and the  $B_{\lambda,j}$  are nonoverlapping open intervals, such that:

$$(P.3) \quad \psi_0(x) = \begin{cases} \zeta(x), & x \in B_\lambda^c \\ \xi(x; \omega_j, \lambda), & x \in B_{\lambda,j}, \end{cases}$$

where  $J(\xi(\cdot; \omega_j, \lambda)) \equiv -\lambda$  on  $B_{\lambda,j}$  for any fixed  $\omega_j$ ;

$$(P.4) \quad f_0(x) = \begin{cases} (1 - \varepsilon)h(x), & x \in B_\lambda^c \\ (1 - \varepsilon)(\sup_{B_{\lambda,j}}(h(x)/k(x; \omega_j, \lambda)))k(x; \omega_j, \lambda), & x \in B_{\lambda,j}, \end{cases}$$

where each  $k$  satisfies  $\xi = -k'/k$  and  $\sup_{B_{\lambda,j}}(h/k)(x)$  is attained at each nonzero, finite endpoint of  $B_{\lambda,j}$ ;

(P.5)  $\int_{B_\lambda} \nu g_0 \, dx \leq 1/2$ , with equality if  $\lambda > 0$ , where  $g_0 = (f_0 - (1 - \varepsilon)h)/\varepsilon$ ;

(P.6)  $A_\lambda = \{x \in (0, \infty) \mid J(\zeta) < -\lambda\} \subseteq B_\lambda$ ;

(P.7)  $\psi_0 \eta(0^+) \nu g_0(0) = 0 \leq \psi_0 \eta(0^+)$

(B) If the pair  $(\psi_0, f_0)$  satisfies (P.1)–(P.7), and if as well  $\psi_0 \in \Psi$ , then the saddlepoint property (2.5) holds in  $\Psi \times \mathcal{F}'_\nu$ .

(C) If the conditions in (B) are satisfied, and  $(\psi_0 \sigma/\nu)'$  is either continuous or nonnegative upper semicontinuous, and vanishes at infinity, then  $(\psi_0, f_0)$  is not only a saddlepoint with respect to  $\mathcal{F}'_\nu$ , but with respect to  $\mathcal{F}_\nu$  as well.

PROOF. The existence and uniqueness of  $F_\nu^0$  follow from Theorem 2, as does part (B). Part (C) may be proven as in Theorem 5 of Huber (1964), by writing  $V(\psi, f)$  as

$$\int_0^\infty (\psi^2 \sigma/\nu) \, dF_\nu / 2 \left( \int_0^\infty ((\psi \sigma)'/\nu) \, dF_\nu \right)^2.$$

To show the sufficiency of (P.1)–(P.7), it suffices to verify (2.6) for the dense subclass of  $\mathcal{G}'$  for which  $\lim_{x \rightarrow \infty} g\nu(x) = 0$ . For this subclass, an integration by parts, using (P.2) and (H.3), establishes the equivalence of (2.6) with

$$(2.7) \quad \int_0^\infty J(\psi_0) \nu (g - g_0) \, dx + 2(\nu g - \nu g_0)(0) (\psi_0 \eta)(0^+) \geq 0.$$

The second term in (2.7) is nonnegative by (P.7). Considering the ranges  $B_\lambda$  and  $B_\lambda^c$  separately, and using (P.3), (P.6) shows that the integral is bounded below by  $-\lambda [\int_0^\infty \nu (g - g_0) \, dx]$ , which is nonnegative by (P.5).

It remains to show that (P.1)–(P.7) are necessary. For this, define

$$(2.8) \quad \frac{\lambda}{2} = \int_{B_\lambda} [2\psi_0 g'_0 + \psi_0^2 g_0] \sigma \, dx,$$

where  $B_\lambda$  is the support, in  $(0, \infty)$ , of  $g_0$ . Represent  $B_\lambda$  as the disjoint union of its maximal components, as in the statement of the theorem. Put  $\xi(x) = \psi_0(x) 1_{B_\lambda}(x)$ , so that  $\psi_0$  agrees with  $\xi$  on  $B_\lambda$ , and with  $\zeta$  on  $B_\lambda^c$ . Condition (P.4) is now an immediate consequence of these definitions and the continuity of  $f_0$ .

It follows from (2.4), the convexity of  $I(F_\nu^t)$ , and (2.8) that

$$(2.9) \quad 0 \leq 2\varepsilon \left\{ \frac{\lambda}{2} - \int_0^\infty [2g'_1 \psi_0 + g_1 \psi_0^2] \sigma \, dx \right\} = \frac{d}{dt} I(F_\nu^t) |_{t=0} < \infty,$$

whenever  $I(F_\nu^t) < \infty$ . This is easily seen to imply that  $\lambda < \infty$ . The choice  $g_1 = g_0/2$  in (2.9) entails  $\lambda \geq 0$ , and the choice  $g_1 = g_0/(2 \int_{B_\lambda} g_0 \nu \, dx)$  then establishes the necessity of (P.5). On the other hand, if  $g_1$  has finite support contained in  $B_\lambda^c$ , the first inequality in (2.9) becomes

$$-\frac{\lambda}{2} \leq - \int_{B_\lambda^c} [2g'_1 \zeta + g_1 \zeta^2] \sigma \, dx = \int_{B_\lambda^c} J(\zeta) \nu g_1 \, dx,$$

after an integration by parts. Thus  $J(\zeta) \geq -\lambda$  on  $B_\lambda^c$ , and (P.6) is established.

Before showing that (P.1)–(P.3) are necessary, we note that they imply that (2.7) holds, hence that (P.7) is necessary as well. To prove that (P.3) is necessary, we first claim that the function

$$\phi(t) = 2(\xi\eta)(t) - \int_0^t \left[ -2(\xi\eta)(x) \frac{\nu'}{\nu}(x) + (\xi^2\eta)(x) \right] dx + \lambda t, \quad t \in B_\lambda$$

is constant on  $B_\lambda$ . It then follows, using (C.1), that  $\xi$  is absolutely continuous there and satisfies  $0 \equiv \phi'(t) = J(\xi) + \lambda$ . To establish the claim, let  $G_\nu$  be any fully stochastic member of  $\mathcal{S}'_\nu$  with support  $[a, b] \subset B_\lambda$ . Using integration by parts to evaluate the second and third terms in  $\int_a^b \phi(t)(g\nu)'(t) dt$ , see that this becomes  $\int_a^b [2\xi g' + \xi^2 g]\sigma dt - \lambda/2$ , which is nonpositive by (2.9). Since  $[a, b]$  is arbitrary,

$$(2.10) \quad \int_{B_\lambda} \phi(g\nu)' dt \leq 0$$

for all fully stochastic  $G_\nu \in \mathcal{S}'_\nu$  with support in  $B_\lambda$ . The claim will follow immediately, once it is shown that equality holds in (2.10) for all such  $G_\nu$ . That equality holds if  $g = g_0$  is (2.8). Here we must approximate  $g_0$  by functions with compact support, and use the fact that  $\lambda = 0$  if  $g_0\nu$  is substochastic.

Suppose, for contradiction, that  $\int_a^b \phi(g_1\nu)' dt < 0$  for some  $[a, b] \subset B_\lambda$  and some  $g_1\nu$  with mass  $1/2$  on  $[a, b]$ . Assume that  $g_0\nu$  has been normed, if necessary, so that it is fully stochastic. Choose  $\alpha \in (0, 1)$  sufficiently small that the function  $g_2 = (g_0 - \alpha g_1)/(1 - \alpha)$  is nonnegative on  $B_\lambda$ . Then by (2.10),

$$0 \geq \int_{B_\lambda} \phi(g_2\nu)' dt = \frac{-\alpha}{1 - \alpha} \int_a^b \phi(g_1\nu)' dt > 0,$$

a contradiction. Thus equality holds in (2.10) and (P.3) is established.

In the presence of piecewise smoothness of  $\psi_0\eta$  ((P.3) and (H.4)), (P.1) is equivalent to continuity and boundedness of  $\psi_0\eta$ . But were these to fail, one could construct sequences  $\{f_n\nu\}$  with  $\sup V(\psi_0, f_n) = \infty$ , a contradiction.

Finally, note that by virtue of (C.2), (C.3) and (P.1), the function  $w(x) = -\log f_0\nu(x)$  has a continuous derivative which is bounded as  $x \rightarrow \infty$ , and  $\int_0^\infty \exp(-w(x)) dx \leq 1/2$ . This implies that  $w(x) \rightarrow \infty$  as  $x \rightarrow \infty$ , which is (P.2).  $\square$

Condition (P.6) of Theorem 3 states that those regions in which  $J(\xi)$  drops below  $-\lambda$  must be contained in the support of  $g_0$ . In Lemma 5 below, we show that every region of support of  $g_0$  contains a subinterval on which  $J(\xi) < -\lambda$ . We require a preliminary result.

**LEMMA 4.** *Let  $K(x)$  be any differentiable function defined for all  $x > a$ , “ $a$ ” arbitrary. Then*

$$\inf_{[a, \infty)} 2K'(x) - K^2(x) \leq 0.$$

*The inequality is strict unless  $K \equiv 0$ .*

**PROOF.** Suppose that the inequality fails, and that  $K \not\equiv 0$ . Then  $K$  increases at least exponentially quickly, and we may assume that  $K(x) > 0$  on  $[a, \infty)$ . It



follows that the function  $u(x) = \exp(-\frac{1}{2} \int_a^x K(z) dz)$  is positive, decreasing, and concave on  $(a, \infty)$ , which is a contradiction.  $\square$

**REMARK 2.1.** From condition (P.6) of Theorem 3,  $B_\lambda = \emptyset$  implies that  $J(\zeta)(x) \geq 0$  for all  $x \geq 0$ . In the univariate location estimation problem (Example 2.1),  $\nu \equiv \eta \equiv \sigma \equiv 1$  and  $J(\zeta) = 2\zeta' - \zeta^2$ , so that  $J(\zeta) \geq 0$  on  $[0, \infty)$  is impossible by Lemma 4. Hence we always have  $B_\lambda \neq \emptyset$  when  $\nu \equiv \zeta \equiv \sigma \equiv 1$ . But in the next section, we will give a special case of the multivariate location estimation problem (Example 3.5) in which  $g_0 \equiv 0$ .

**LEMMA 5.** *If  $B_\lambda \neq \emptyset$ , then  $A_\lambda \cap B_{\lambda,j} \neq \emptyset$  for all  $j \leq N(\lambda)$ .*

**PROOF.** Let  $(c, d) \subset B_{\lambda,j} = (a, b)$  be arbitrary. If  $A_\lambda \cap B_{\lambda,j} = \emptyset$ , then

$$0 \leq \int_c^d [J(\zeta) - J(\xi)]\nu f_0 dx = 2(\zeta - \xi)\sigma f_0 \Big|_c^d - \int_c^d (\zeta - \xi)^2 \sigma f_0 dx;$$

so that by the Cauchy-Schwarz inequality,

$$(2.11) \quad 0 \leq \left[ \int_c^d (\zeta - \xi)\sigma f_0 dx \right]^2 \leq \int_c^d (\zeta - \xi)^2 \sigma f_0 dx \leq 2(\zeta - \xi)\sigma f_0 \Big|_c^d.$$

Let  $d \rightarrow b$  in (2.11). If  $b < \infty$ , then  $(\zeta - \xi)(b) = 0$  and (2.11) implies that  $\zeta < \xi$  throughout  $B_{\lambda,j}$ . If  $b = \infty$ , this is implied by Lemma 4, with  $K(d) = \int_c^d (\zeta - \xi)\sigma f_0 dx$ . Now integrate “ $\zeta < \xi$ ” over  $(a, c)$  to obtain the contradiction that  $g_0(c) < 0$ , if  $g_0(a) = 0$ . If  $g_0(a) > 0$ , then  $a = 0$  and (H.5) is contradicted, using (P.7).  $\square$

Suppose that under the conditions of Theorem 3, the pair  $(\psi_0, f_0)$  satisfies conditions (P.1)–(P.7), where the set  $B_\lambda$  (the support of  $g_0$ ) is nonempty and where  $A_\lambda = \{x \in (0, \infty) \mid J(\zeta) < -\lambda\}$  is a single interval  $(c_\lambda, d_\lambda)$ . By condition (P.6) and Lemma 5,  $B_\lambda$  must be a single interval  $(a_\lambda, b_\lambda) \supseteq (c_\lambda, d_\lambda)$ . If  $a_\lambda \neq 0$  ( $b_\lambda \neq \infty$ ), it must be the largest (smallest) zero of  $\xi - \zeta$  to the left (right) of  $c_\lambda(d_\lambda)$ , since extending the support of  $g_0$  beyond this point would contradict Lemma 5. Thus in the special case where  $A_\lambda$  is a single interval, the solution must have the form stated in Theorem 6 below. The problem is then reduced to a numerical one of determining four constants  $(\omega, \bar{\lambda}, a_{\bar{\lambda}}, b_{\bar{\lambda}})$ . We have:

**THEOREM 6.** *Suppose that for all  $\lambda > 0$ ,  $A_\lambda$  is a single interval  $(c_\lambda, d_\lambda)$ . Then there exists a unique pair  $(\omega, \bar{\lambda})$ , where  $\omega \in [-\infty, \infty]$  and  $0 \leq \bar{\lambda} \leq \inf\{\lambda \mid A_\lambda = \emptyset\}$ , such that the pair  $(\psi_0, f_0)$ , defined below, satisfies the given conditions.*

$$(i) \quad \psi_0(x) = \begin{cases} \zeta(x), & x \in (0, a_{\bar{\lambda}}) \cup [b_{\bar{\lambda}}, \infty), \\ \xi(x; \omega, \bar{\lambda}), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]; \end{cases}$$

where  $J(\xi) \equiv -\bar{\lambda}$ ,

$$a_{\bar{\lambda}} = \sup\{x \leq c_{\bar{\lambda}} \mid (\xi - \zeta)(x) = 0\}, \quad (\text{where } \sup \emptyset = 0),$$

$$b_{\bar{\lambda}} = \inf\{x \geq d_{\bar{\lambda}} \mid (\xi - \zeta)(x) = 0\}, \quad (\text{where } \inf \emptyset = \infty).$$

$$(ii) \quad f_0(x) = \begin{cases} (1 - \varepsilon)h(x), & x \in (0, a_{\bar{\lambda}}] \cup [b_{\bar{\lambda}}, \infty) \\ (1 - \varepsilon)sk(x), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]; \end{cases}$$

where  $\xi = -k'/k$  and

$$s = \sup_{[a_{\bar{\lambda}}, b_{\bar{\lambda}}]} (h/k)(x)$$

$$= \begin{cases} (h/k)(a_{\bar{\lambda}}) & \text{if } b_{\bar{\lambda}} = \infty, \\ (h/k)(b_{\bar{\lambda}}) & \text{if } a_{\bar{\lambda}} = 0, \\ (h/k)(a_{\bar{\lambda}}) = (h/k)(b_{\bar{\lambda}}) & \text{if } 0 < a_{\bar{\lambda}} < b_{\bar{\lambda}} < \infty. \end{cases}$$

$$(iii) \quad \frac{1 - \varepsilon}{\varepsilon} \int_{a_{\bar{\lambda}}}^{b_{\bar{\lambda}}} \nu(x)[sk(x) - h(x)] dx \leq \frac{1}{2}, \quad \text{with equality if } \bar{\lambda} > 0.$$

If  $\psi_0 \in \Psi$ , then  $(\psi_0, f_0)$  possesses the saddlepoint property:

$$V(\psi_0, f) = (1/I(F_\nu^0)) = V(\psi_0, f_0) \leq V(\psi, f_0)$$

for all  $\psi \in \Psi$  and  $F \in \mathcal{F}$ .

**3. Examples.** We first consider special cases of Example 2.1 of the previous section, that is, the case  $\eta = \nu = \sigma \equiv 1$  corresponding to the one-dimensional location parameter estimation problem. We first note that the solution to  $J(\xi) = 2\xi' - \xi^2 \equiv -\lambda$ , where  $\lambda \geq 0$ , must have one of the following three forms on each of its intervals of support:

$$(3.1) \quad \xi(x; \omega, \lambda) = \sqrt{\lambda} \tanh[-(\sqrt{\lambda}/2)(x - \omega)] \quad (\text{decreasing in } x); \text{ or}$$

$$(3.2) \quad \xi(x; \omega, \lambda) = \sqrt{\lambda} \quad (\text{constant}); \text{ or}$$

$$(3.3) \quad \xi(x; \omega, \lambda) = \sqrt{\lambda} \coth[-(\sqrt{\lambda}/2)(x - \omega)] \quad (\text{increasing in } x).$$

The next four examples will give the minimax solution  $(\psi_0, f_0)$ , where  $f_0 = (1 - \varepsilon)h + \varepsilon g_0$  and  $\psi_0 = -f_0'/f_0$ , corresponding to four choices of  $h$ . In each case only the restriction of the solution to the set  $[0, \infty)$  will be given (extension to the context of Example 2.1 is given by symmetry:  $f_0(-x) = f_0(x)$  for all  $x$ ,  $\psi_0(-x) = -\psi_0(x)$  for all  $x \neq 0$ ). In Example 3.1 the support of  $g_0$  is  $[a, \infty)$  for some  $a > 0$ ; in Example 3.2,  $\text{supp}\{g_0\} = [a, b]$  for some  $0 < a < b < \infty$ ; in Example 3.3,  $\text{supp}\{g_0\} = [0, b]$  for some  $b > 0$ ; and in Example 3.4,  $\text{supp}\{g_0\} = \cup_{i=0}^\infty [a_i, b_i]$  where the  $[a_i, b_i]$  are a countable collection of nonempty disjoint sets.

**EXAMPLE 3.1.** Let  $\nu = \eta = \sigma \equiv 1$  and assume that  $\zeta$  is increasing (i.e.,  $h$  is strongly unimodal). Then it follows easily from Theorem 6 that  $B_\lambda$  must be a single half-infinite interval. Since it is easily seen that solutions to  $J(\zeta) = -\lambda$

of form (3.1) or (3.3) are impossible on such  $B_\lambda$ , only the constant solution (3.2) is left, yielding Huber's (1964) result that  $\psi_0 = -f'_0/f_0$  is given by (1.2). We note that this example contains as special cases densities of the form  $h(x) = \text{const} \cdot e^{-|x|^{1+\alpha}}$ ,  $\alpha > 0$ . In the limiting case that  $h$  is the Laplace density  $\frac{1}{2} e^{-|x|}$ , the solution is  $\psi_0(x) = (1 - \epsilon)$  for  $x > 0$ , and  $f_0(x) = \frac{1}{2}(1 - \epsilon)e^{-(1-\epsilon)x}$  for  $x > 0$ .

**EXAMPLE 3.2.** Let  $\nu = \eta = \sigma \equiv 1$  and let  $h$  be the Cauchy density

$$h(x) = (1/\pi)(1/(1 + x^2)).$$

In this case,

$$\zeta(x) = 2x/(1 + x^2), \quad J(\zeta) = 4(1 - 2x^2)/(1 + x^2)^2,$$

and

$$A_\lambda = \{x \in (0, \infty) \mid J(\zeta) < -\lambda\} = \begin{cases} (c_\lambda, d_\lambda), & \lambda < 4/3, \\ \emptyset, & \lambda \geq 4/3; \end{cases}$$

where  $c_\lambda^2 = (4 - \lambda - 2\sqrt{4 - 3\lambda})/\lambda$ ,  $d_\lambda^2 = (4 - \lambda + 2\sqrt{4 - 3\lambda})/\lambda$ . In this case, Theorem 6 applies, so that the support of  $g_0$  is a set  $(a_{\bar{\lambda}}, b_{\bar{\lambda}})$  (as defined in the statement of Theorem 6) on which  $J(\xi) \equiv -\bar{\lambda}$ . Now if  $\xi$  were the constant solution (3.2) to  $J(\xi) \equiv -\bar{\lambda}$  on  $(a_{\bar{\lambda}}, b_{\bar{\lambda}})$ , the continuity of  $\psi_0$  would force  $b_{\bar{\lambda}} > 1$  so that  $\psi'_0(b_+) < 0 = \psi'_0(b_-)$ , which violates (P.6). Similarly the "coth" solution (3.3) to  $J(\xi) \equiv -\bar{\lambda}$  also violates (P.6). This leaves only the "tanh" solution (3.1). For  $\lambda > 0$ ,  $\xi - \zeta$  has three zeros  $a_\lambda, e_\lambda, b_\lambda$ ; with  $a_\lambda (> 0), b_\lambda (< \infty) \in A_\lambda^c$  and  $e_\lambda \in A_\lambda$  (to satisfy (P.6)). Thus, for  $\lambda \in (0, 4/3)$  and  $e_\lambda^2 \in ((4 - \lambda - 2\sqrt{4 - 3\lambda})/\lambda, (4 - \lambda + 2\sqrt{4 - 3\lambda})/\lambda)$ , set  $\omega = e_\lambda + (2/\sqrt{\lambda})\tanh^{-1}(2e_\lambda/\sqrt{\lambda}(1 + e_\lambda^2))$ , so that  $\xi(e_\lambda; \lambda, \omega) = \zeta(e_\lambda)$ . Let  $a_\lambda$  and  $b_\lambda$  be the other two zeros of  $\xi - \zeta$ . Put  $k(x; \omega, \lambda) = \cosh^2((-\sqrt{\lambda}/2)(x - \omega))$  so that  $-k'/k = \xi$ . Then by Theorem 6 there exists a unique pair  $(\bar{\lambda}, e_{\bar{\lambda}})$  in the indicated region satisfying

$$s = (h/k)(a_{\bar{\lambda}}) = (h/k)(b_{\bar{\lambda}}) = \sup_{(a_{\bar{\lambda}}, b_{\bar{\lambda}})} (h/k)(x),$$

and

$$\frac{1 - \epsilon}{\epsilon} \int_{a_{\bar{\lambda}}}^{b_{\bar{\lambda}}} [sk(x) - h(x)] dx = \frac{1}{2};$$

and the optimal pair is given by

$$\begin{aligned} \psi_0(x) &= 2x/(1 + x^2) & x \notin [a_{\bar{\lambda}}, b_{\bar{\lambda}}] \\ &= \sqrt{\bar{\lambda}} \tanh(-(\sqrt{\bar{\lambda}}/2)(x - \omega)), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]; \end{aligned}$$

and

$$\begin{aligned} f_0(x) &= (1 - \epsilon)/(\pi(1 + x^2)) & x \notin [a_{\bar{\lambda}}, b_{\bar{\lambda}}], \\ &= (1 - \epsilon)s \cdot \cosh^2((-\sqrt{\bar{\lambda}}/2)(x - \omega)), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]. \end{aligned}$$

Here the five constants  $(\lambda, \omega, s, a, b)$  are determined by the five side conditions that both  $\psi_0$  and  $f_0$  are continuous at both  $a$  and  $b$  and that  $\int_0^\infty f_0(x) dx = 1/2$ .

Figures 1, 2, 3 give plots of  $h$ ,  $\zeta$  and  $J(\zeta)$  (dotted curves) along with plots of  $f_0$ ,  $\psi_0$  and  $J(\psi_0)$ , respectively (solid curves) when  $h$  is the Cauchy density. For a large class of densities for which the minimax solution is qualitatively similar to the solution in the Cauchy case, consider the class of  $t$ -densities

$$h_k(x) = (\text{const}) \cdot (1 + x^2)^{-(1+k)/2}.$$

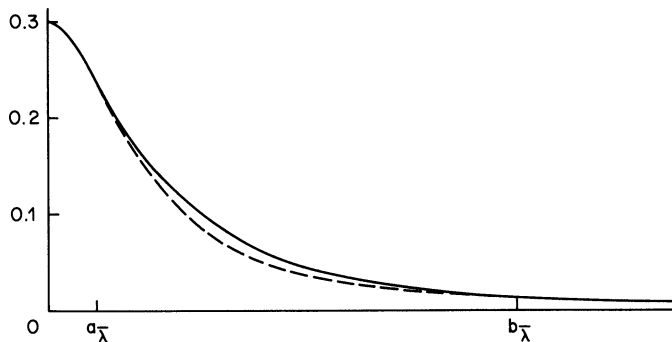


FIG. 1. Cauchy density  $(1 - \epsilon)h_0$  (---) and least favorable  $f_0$  (—).

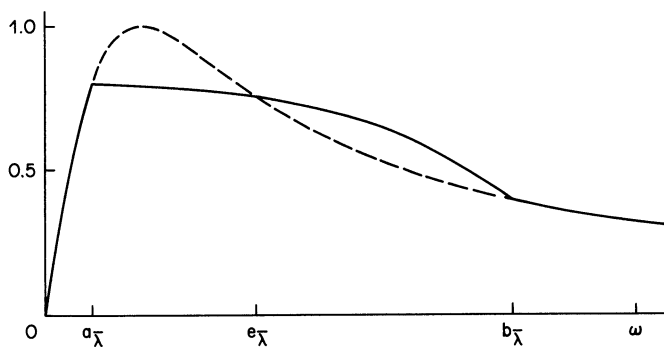


FIG. 2.  $\zeta = -h'/h$  (---) and  $\psi_0 = -f'_0/f_0$  (—).

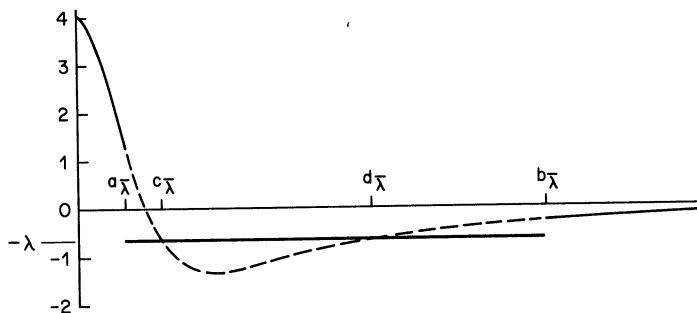


FIG. 3.  $J(\zeta)$  (---) and  $J(\psi_0)$  (—).

Then we have that  $\zeta(x) = (k + 1)x/(1 + x^2)$  and

$$J(\zeta)(x) = ((k + 1)/(1 + x^2)^2)(2 - (k + 3)x^2),$$

so it is clear that the minimax solution is similar to that of the Cauchy case (which is the special case  $k = 1$ ).

**EXAMPLE 3.3.** Let  $\nu = \eta = \sigma \equiv 1$  and let  $h$  be the density

$$h(x) = c_\alpha e^{-|x|^{1-\alpha}},$$

where  $0 < \alpha < 1/2$  and  $c_\alpha$  is chosen so that  $\int_0^\infty h(x) dx = 1/2$ . (Note that if  $\alpha \geq 1/2$ , then  $I(H_\nu) = \infty$ .) Then we have that  $\zeta(x) = -h'(x)/h(x) = (1 - \alpha)x^{-\alpha}$  and

$$J(\zeta)(x) = 2\zeta'(x) - \zeta^2(x) = -((1 - \alpha)/x^{1+\alpha})[2\alpha + (1 - \alpha)x^{1-\alpha}] < 0$$

for all  $x > 0$ . Note that  $J(\zeta)(x)$  is strictly monotone increasing from  $-\infty$  at  $0^+$  to  $0$  at  $\infty$ . So we have

$$\begin{aligned} A_\lambda &= \{x \mid J(\zeta)(x) < -\lambda\} \\ &= (0, d), \quad \text{where } -(1 - \alpha)d^{-1-\alpha}[2\alpha + (1 - \alpha)d^{1-\alpha}] = -\lambda. \end{aligned}$$

It follows from Theorem 6 that  $B_\lambda = \text{supp}(g_0)$  is of the form  $(0, b)$  for some  $b > d$ . It is easy to see that both the constant solution (3.2) and the ‘‘coth’’ solution (3.3) to the equation  $J(\xi) \equiv -\lambda$  on  $(0, b)$  would violate the necessary condition that  $A_\lambda \subseteq B_\lambda$ , and so these solutions are impossible. So the solution has the form

$$\begin{aligned} \psi_0(x) &= \lambda^{1/2} \tanh((-\sqrt{\lambda}/2)(x - \omega)), & 0 \leq x \leq b \\ &= \zeta(x) = (1 - \alpha)/x^\alpha & x > b; \end{aligned}$$

and

$$\begin{aligned} f_0(x) &= (1 - \epsilon)h(b) \frac{\cosh^2((-\sqrt{\lambda}/2)(x - \omega))}{\cosh^2((\sqrt{\lambda}/2)(b - \omega))} & 0 \leq x \leq b \\ &= (1 - \epsilon)h(x) & x \geq b. \end{aligned}$$

The constants are determined by (i) continuity of  $\psi_0$  at  $b$ , (ii) the condition  $g_0(0) = 0$  (by condition (P.7) since  $\psi_0(0^+) > 0$ ), and (iii)  $\int_0^\infty f_0 dx = 1/2$ .

**EXAMPLE 3.4.** Let  $\nu = \eta = \sigma \equiv 1$  and let  $h$  be the density with corresponding  $\zeta = -h'/h$  given by

$$\begin{aligned} \zeta(x) &= x, & 0 \leq x \leq 2 \\ &= 2 + \sin(x - 2), & x > 2. \end{aligned}$$

Then it is easy to see that the set of points at which  $J(\zeta) = 2\zeta' - \zeta^2$  attains its minimum over  $\mathbb{R}^+$  is  $\{x: x = x_0 + 2\pi k, k = 0, 1, 2, \dots\}$ , where  $x_0$  is a point in  $[0, 2 + 2\pi]$ . Since  $B_\lambda$  is nonempty by Remark 2.1 and  $A_\lambda \cap B_{\lambda_j} \neq \emptyset$  for all  $j \leq N(\lambda)$  by Lemma 5, it follows that  $A_\lambda$  contains an interval in the neighborhood

of each of the points of  $\{x_0 + 2\pi k, k = 0, 1, \dots\}$ . For sufficiently small values of  $\epsilon$ , one can easily verify that  $\psi_0 = -f'_0/f_0$  has the following form:

$$\begin{aligned} \psi_0(x) &= \zeta(x), & x \in [0, 2 + 2\pi] \setminus [a, b] \\ &= \sqrt{\lambda} \tanh[(\sqrt{\lambda}/2)(\omega - x)], & x \in [a, b] \\ &= \psi_0(x - 2k\pi), & x \in [2 + 2k\pi, 2 + 2(k + 1)\pi], \quad k = 1, 2, \dots, \end{aligned}$$

where  $2 < a < x_0 < b < a + 2\pi$ . Here the constants  $(a, b, \omega, \lambda)$  are determined by the side conditions that both  $\psi_0$  and  $f_0$  are continuous and that  $g_0 = [f_0 - (1 - \epsilon)h]/\epsilon$  (which must necessarily be positive on its support  $B_\lambda = \cup_{k=0}^\infty (a + 2\pi k, b + 2\pi k)$ ) satisfies  $\int_0^\infty g_0 dx = 1/2$ .

The  $\zeta$  of Example 3.4, chosen for its mathematical convenience, is of no practical interest. A qualitatively similar result is obtained for any choice of  $\zeta$  for which  $J(\zeta(x))$  attains its minimum infinitely often as  $x \rightarrow \infty$ .

The next two examples are special cases of Example 2.2 where  $\nu(x) = 1/2mC_mx^{m-1}$  and  $\eta_i(x) = (2/(m + 2))^i(x^{2i}/m)$  corresponding to asymptotic variance functionals for M-estimators of location ( $i = 0$ ) and scale ( $i = 1$ ) in affinely invariant  $m$ -dimensional models (see Huber 1977, 1981). Unlike the case  $\eta = \nu \equiv 1$  (where Remark 2.1 applies), one can find choices of  $h$  for which the least informative  $f_0 = (1 - \epsilon)h + \epsilon g_0$  is substochastic. This is illustrated for  $m$ -dimensional location ( $i = 0$ ) and scale ( $i = 1$ ) in Examples 3.5 and 3.6, respectively.

**EXAMPLE 3.5.** Let  $m \geq 2$ ,  $\nu(x) = 1/2mC_mx^{m-1}$  and  $\eta(x) = 1/m$ . Then  $\sigma = \eta\nu$  satisfies  $(\sigma'/\sigma)(x) = (m - 1)/x$ . Now let

$$h(x) = \text{const} \cdot (1 + x^2)^{-(m+k)/2}, \quad x > 0$$

so that we have

$$(h\nu)(x) = \text{const} \cdot x^{m-1}(1 + x^2)^{-(m+k)/2}, \quad x > 0.$$

The minimax problem is the multivariate location estimation problem with  $h\nu$  an  $m$ -dimensional "Student"  $t$ -density. Then we have  $\zeta(x) = (m + k)x/(1 + x^2)$ ,  $\zeta'(x) = (m + k)(1 - x^2)/[(1 + x^2)^2]$ , and

$$\begin{aligned} J(\zeta)(x) &= (2\zeta' - \zeta^2 + 2\zeta(\sigma'/\sigma))\eta(x) \\ &= ((m + k)/m(1 + x^2)^2)[2m + (m - 4 - k)x^2]. \end{aligned}$$

Now suppose that  $m \geq k + 4$ . Then  $J(\zeta)(x) > 0$  for all  $x > 0$ , and so the choices  $\lambda = 0$ ,  $B_\lambda = \emptyset$  and  $g_0 \equiv 0$  yield the minimax solution by Theorem 3 and Lemma 5. That is, the least informative distribution has the substochastic density  $f_0 = (1 - \epsilon)h$ , and the minimax  $\psi_0 \equiv \zeta$ .

Note that the case  $m = 1$  yields the univariate  $t$ -densities which were considered in Example 3.2.

**EXAMPLE 3.6.** Let  $m \geq 2$ ,  $\nu(x) = 1/2mC_mx^{m-1}$ , and  $\eta(x) = 2x^2/[m(m + 2)]$ . Then  $\sigma = \eta\nu$  satisfies  $(\sigma'/\sigma)(x) = (m + 1)/x$ . Consider the case of the normal

density; i.e., set

$$h(x) = (2\pi)^{-m/2} \exp(-x^2/2).$$

Then in the minimax solution, the support of  $g_0 = [f_0 - (1 - \epsilon)h]/\epsilon$  must be the set  $B_\lambda$  on which

$$\begin{aligned} J(\xi)(x) &= (2\xi' - \xi^2 + 2\xi(\sigma'/\sigma))\eta(x) \\ &= [2x^2\xi'(x) - x^2\xi^2(x) + 2(m + 1)x\xi(x)]/(2/m(m + 1)) \equiv -\lambda. \end{aligned}$$

Recall from the proof of Theorem 3 that we must have  $\lambda \geq 0$  and that  $g_0$  can be substochastic only if  $\lambda = 0$ . One can then easily see that the minimax solution always has the form

$$(3.4) \quad \begin{aligned} \psi_0(x) &= x, & x \leq b \\ &= b^2/x, & x \geq b; \end{aligned}$$

and

$$(3.5) \quad \begin{aligned} f_0(x) &= (1 - \epsilon)h(x), & x \leq b \\ &= (1 - \epsilon)h(b)(b/x)^{b^2}, & x \geq b \end{aligned}$$

Note that  $\lambda = 2b^2(b^2 - 2m)/[m(m + 2)]$ . Define  $\epsilon(m)$  by  $(1 - \epsilon(m))^{-1} = 4\chi_m^2(2m) + X_m^2(2m)$ , where  $\chi_m^2$  and  $X_m^2$  denote the density and distribution function, respectively, of a chi-squared random variable with  $m$  degrees of freedom. Note that  $\epsilon(2) = (1 + e^2)^{-1} \doteq .119$  and  $\epsilon(m) \rightarrow 0$  as  $m \rightarrow \infty$ . In the case where  $\epsilon < \epsilon(m)$ , the equation  $\int_0^\infty g_0(x) dx = 1/2$ , or equivalently

$$(3.6) \quad 1/(1 - \epsilon) = (2b^2\chi_m^2(b^2)/(b^2 - m)) + X_m^2(b^2),$$

can be seen to have a solution  $b \geq \sqrt{2m}$  so that  $\lambda > 0$ . In the case  $\epsilon > \epsilon(m)$ , we must have  $\lambda = 0$ , so that  $b = \sqrt{2m}$  and the resulting  $f_0$  is substochastic.

Huber (1977, 1981) considers the same minimax problem with a side condition restricting the distributions under consideration to be *proper*. The solution to Huber's problem has form (3.4), (3.5) when  $\epsilon < \epsilon(m)$  but not when  $\epsilon > \epsilon(m)$ . In the latter case, the least informative *proper* density of form  $(1 - \epsilon)h + \epsilon g$  places contaminating mass on sets of form  $[0, a]$  as well as  $[b, \infty)$  (see pages 232–236 of Huber, 1981, for details). We remark that if a restriction to proper distributions is added to the hypothesis of our Theorem 3, then the theorem goes through with the only change being that  $\lambda$  need not satisfy  $\lambda \geq 0$ . In Huber's solution, it is easily checked that  $\lambda < 0$  when  $\epsilon > \epsilon(m)$ .

**4. Theory for redescending M-estimators.** In Section 2 a theory was presented for finding minimax pairs  $(\psi_0, f_0)$  in  $\Psi \times \mathcal{F}$ . Let  $r, 0 < r < \infty$ , be fixed and define  $\Psi_r$  to be the subclass of continuous  $\psi$ 's in  $\Psi$  which satisfy  $\psi(x) = 0$  for all  $x \geq r$ . The class  $\Psi_r$  of "redescenders" is useful for obtaining consistent M-estimators when the unknown error distributions are assumed to be symmetric in a central region and asymmetric in the tail regions. For proofs of consistency

and asymptotic normality, see Collins (1976, 1982) for the scale-known location case, and see Wiens (1982) and Wiens and Zheng (1983) for the case of estimating both location and scale.

The theory of Section 2 essentially goes through with  $\Psi$  replaced by  $\Psi_r$  with one important exception: the side condition  $\Psi_0(r) = 0$  forces  $f'_0(r) = 0$  so that  $B_\lambda$ , the support of  $g_0$ , must contain a neighborhood of  $r$ . But one can find examples in which no pair  $(\psi_0, f_0)$  satisfying the necessary condition (P.6) of Theorem 3 (that  $A_\lambda = \{x \mid J(\zeta)(x) < -\lambda\} \subseteq B_\lambda$ ) can possibly satisfy the condition  $r \in B_\lambda$ . Such an example—for which the variance functional fails to have a saddlepoint in  $\Psi_r \times \mathcal{F}$ —is given in Wiens (1982). To make sure that there is a minimax  $\psi_0$  in  $\Psi_r$ , we will restrict our attention to densities  $h$  for which the condition  $A_\lambda \subseteq B_\lambda$  forces  $r \in B_\lambda$ .

We assume

- (B1)  $\zeta$  is continuously differentiable and positive on  $(0, r)$ , and bounded on  $[0, r]$ . Let  $\xi(x; \omega, \lambda) = \xi(x, \lambda)$  be the solution to  $J(\xi) \equiv -\lambda \in [0, \infty]$ , passing through  $(r, 0)$ . Assume that  $\sigma$  and  $\eta$  are such that the following condition is satisfied.
- (B2) For fixed  $\lambda > 0$ ,  $\xi(x, \lambda)$  is strictly decreasing in  $x \in (0, r)$ .

The following lemma gives an easily checked condition ensuring that (B2) is satisfied, and gives some further properties of  $\xi(x, \lambda)$ .

**LEMMA 7.** *For fixed  $x \in (0, r)$ ,  $\xi(x, \lambda)$  is a continuously differentiable function of  $\lambda$ , with*

- i)  $\xi(x, \lambda) > 0$  ( $\lambda > 0$ ),    ii)  $(d/d\lambda)\xi(x, \lambda) > 0$ ,
- iii)  $\lim_{\lambda \rightarrow 0} \xi(x, \lambda) = 0$ ,    iv)  $\lim_{\lambda \rightarrow \infty} \xi(x, \lambda) = \infty$ .

Furthermore, assumption (B2) is implied by the following condition.

(B2') *For fixed  $\lambda > 0$ ,  $(\sigma'/\sigma)(x) + ((\sigma'/\sigma)(x) + (\lambda/\eta(x)))^{1/2}$  is strictly decreasing in  $x \in (0, r)$ .*

*In particular, if  $(\sigma'/\sigma)(x)$  is nonnegative and decreasing, and  $\eta(x)$  is nondecreasing, then (B2) holds.*

**PROOF.** See the proof of Lemma 4.5 of Wiens (1982).  $\square$

Define, for  $\lambda \geq 0$ ,

$$A_\lambda = \{x \in (0, r) \mid J(\zeta) < -\lambda\},$$

$$B_\lambda = \{x \in (0, r) \mid \xi(x, \lambda) < \zeta(x)\},$$

and note that  $\{A_\lambda\}$  is nonincreasing, and that  $\{B_\lambda\}$  is strictly decreasing from  $(0, r)$  to  $\emptyset$  as  $\lambda$  increases from 0 to  $\infty$ .

Under certain condition on  $\zeta$ , condition (P.6) of Theorem 3 is satisfied by any pair  $(A_\lambda, B_\lambda)$ .



LEMMA 8. *If either of the following conditions holds:*

- (a)  $\zeta$  is nondecreasing on  $(0, r)$ ,
- (b)  $J(\zeta)$  is nonincreasing on  $A_0$ ;

then  $A_\lambda \subseteq B_\lambda$  for all  $\lambda \geq 0$  and  $B_\lambda$  is a single interval  $(a_\lambda, r)$ .

PROOF. By (iii) of Lemma 7,  $B_\lambda$  contains an interval  $(a_\lambda, r)$  for all  $\lambda$ , and  $a_\lambda \downarrow 0$  as  $\lambda \downarrow 0$ . Let  $\{(a_\lambda^j, b_\lambda^j), 0 \leq j < N(\lambda)\}$  be the remaining components of  $B_\lambda$ . Put  $\lambda_0 = \inf\{\lambda \mid A_\lambda \setminus B_\lambda \neq \emptyset\}$ . It suffices to show that  $N(\lambda) = 0$  for  $\lambda < \lambda_0$ , and that  $\lambda_0 = \infty$ .

To prove the first statement, assume without loss of generality that  $\lambda_0 > 0$ , let  $\lambda \in (0, \lambda_0)$ , and choose  $j < N(\lambda)$  if  $N(\lambda) > 0$ . Then  $\xi(b_\lambda^j, \lambda) = \zeta(b_\lambda^j)$ , but

$$(4.1) \quad b_\lambda^j \notin A_\lambda = \{x \mid 2((\xi - \zeta)\sigma)' > (\xi^2 - \zeta^2)\sigma\}.$$

Thus  $(\xi - \zeta)'(b_\lambda^j) \leq 0$ . Strict inequality contradicts  $\emptyset \neq (a_\lambda^j, b_\lambda^j) \subseteq B_\lambda$ , so we have that  $(\xi - \zeta)(b_\lambda^j) = (\xi - \zeta)'(b_\lambda^j) = 0$  for all  $\lambda < \lambda_0$ . The implicit function theorem then implies that  $\lambda'(b^j) \equiv 0$  for  $\lambda \in (0, \lambda_0)$ , contradicting the fact that  $b_\lambda^j$  is an invertible function of  $\lambda$ .

Now assume that condition (b) holds, so that  $A_\lambda$  is a single interval  $(c_\lambda, r)$  for all  $\lambda$ . Suppose that  $\lambda_0 < \lambda < \infty$ , and let  $x \in A_\lambda \setminus B_\lambda$ . As at (4.1),  $(\xi - \zeta)\sigma$  is increasing and positive at  $x$ , and hence must possess a stationary point  $y \in (x, r)$  at which it is positive. But then  $y \notin A_\lambda$ , a contradiction.

Under condition (a),  $N(\lambda) = 0$  for all  $\lambda$ . If  $\lambda_0 < \infty$ , it is easy to see that for some  $\lambda$ ,  $a_\lambda$  is a boundary point of both  $A_\lambda$  and  $B_\lambda$ . But this contradicts (a), using (4.1) again.  $\square$

The statistical problem of minimaxing  $V(\psi, f)$  over  $\Psi_r \times \mathcal{F}$  is meaningless unless the least informative member  $F_0$  has positive information. Suppose that

$$I_r(F_0^0) = 2 \int_0^r \left(\frac{f_0'}{f_0}\right)^2 f_0 \sigma \, dx = 0.$$

Then on  $[0, r]$ ,  $f_0' \equiv 0$ , so that  $f_0(x) \geq (1 - \epsilon)h(0)$ , and

$$\int_0^r \nu g_0 \, dx \geq \frac{1 - \epsilon}{\epsilon} \int_0^r \nu(x)[h(0) - h(x)] \, dx.$$

If  $\epsilon$  is sufficiently small, then this last term exceeds  $1/2$ , contradicting  $g_0 \in \mathcal{G}$ . This bound on  $\epsilon$ , say  $\epsilon^*$ , is clearly also necessary, and so  $I_r(F_0^0) > 0$  iff  $\epsilon < \epsilon^*$ . Although  $\psi_0 = -f_0'/f_0$  will not, in general, belong to  $\Psi_r$ , it is easy to construct a sequence  $\{\psi_n\} \subseteq \Psi_r$ , with  $\psi_n \rightarrow \psi_0$  pointwise on  $(-r, r)$  such that

$$V(\psi_n, f_0) \rightarrow V(\psi_0, f_0) = 1/I_r(F_0^0) < \infty.$$

We thus have

LEMMA 9. *In order that  $\inf_{\Psi} \sup_{\mathcal{F}} V(\psi, f)$  be finite, it is necessary and sufficient that  $\epsilon$  satisfy  $\epsilon < \epsilon^*$ , where*

$$\frac{1 - \epsilon^*}{\epsilon^*} \int_0^r \nu(x)[h(0) - h(x)] \, dx = \frac{1}{2}.$$

Note that  $\epsilon^* \rightarrow 1$  as  $r \rightarrow \infty$ . Now assume that either condition of Lemma 8 holds. Let  $k(x, \lambda)$  satisfy  $\xi = -k'/k$ . Define

$$(4.2) \quad \psi(x, \lambda) = \{\zeta(x), \xi(x, \lambda), 0\}$$

$$(4.3) \quad f(x, \lambda) = \{(1 - \epsilon)h(x), (1 - \epsilon)h(a_\lambda)k(x, \lambda)/(k(a_\lambda, \lambda)), 0\}$$

on  $[0, a_\lambda]$ ,  $[a_\lambda, r]$ ,  $[r, \infty)$ , respectively. Then conditions (P.1), (P.2), (P.3), (P.6) and (P.7) of Theorem 3 are satisfied. Condition (P.4) follows from the fact that  $\xi - \zeta = (\log h/k)'$  is decreasing through zero at  $a_\lambda$ . For (P.5), define

$$\alpha(\lambda) = \frac{1 - \epsilon}{\epsilon} \int_{a_\lambda}^r \nu(x) \left[ \frac{h(a_\lambda)}{k(a_\lambda, \lambda)} k(x, \lambda) - h(x) \right] dx - \frac{1}{2}.$$

Using (B.1) and Lemma 7, verify that if  $\epsilon < \epsilon^*$ , then  $\alpha(\lambda)$  is strictly decreasing, from positive to negative values, and hence has a unique zero  $\bar{\lambda}$ . This is (P.5).

The conditions of Theorem 3 having been met, we now have:

**THEOREM 10.** *Assume that  $\epsilon < \epsilon^*$ , and that (B1), (B2) and either condition of Lemma 8 hold. Let  $(\bar{\lambda}, a_{\bar{\lambda}}) \in (0, \infty) \times (0, r)$  be the unique pair satisfying  $\alpha(\bar{\lambda}) = 0$ ,  $\xi(a_{\bar{\lambda}}, \bar{\lambda}) = \zeta(a_{\bar{\lambda}})$ . Put  $\psi_0(x) = \psi(x, \bar{\lambda})$ ,  $f_0(x) = f(x, \bar{\lambda})$  in (4.2) and (4.3). Then*

$$V(\psi_0, f) \leq 1/I_r(F_r^0) = V(\psi_0, f_0) \leq V(\psi, f_0)$$

for all  $\psi \in \Psi_r$  and  $f \in \mathcal{F}$ .

It can be verified that under the conditions of Theorem 10, the minimax solutions to the multivariate location and scale problems of Example 2.2 are of the form (4.2), (4.3). Details of the solutions and examples are found in Wiens and Zheng (1983).

### REFERENCES

BICKEL, P. J. (1984). Robust regression based on infinitesimal neighborhoods. *Ann. Statist.* **12** 1349-1368.  
 COLLINS, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.* **4** 68-85.  
 COLLINS, J. R. (1982). Robust M-estimators of location vectors. *J. Multivariate Anal.* **12** 480-492.  
 DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*. (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth, Belmont, CA.  
 HAMPEL, F. R., ROUSSEEUW, P. J. and RONCHETTI, E. (1981). The change-of-variance curve and optimal redescending M-estimators. *J. Amer. Statist. Assoc.* **76** 643-648.  
 HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.  
 HUBER, P. J. (1977). Robust covariances. In *Statistical Decision Theory and Related Topics, II*. (S. S. Gupta and D. S. Moore, eds.) Academic, New York.  
 HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.  
 HUBER, P. J. (1983). Minimax aspects of bounded-influence regression. *J. Amer. Statist. Assoc.* **78** 66-80.  
 JOINER, B. L. and HALL, D. J. (1980). The ubiquitous role of  $f'/f$  in efficiency robust estimation of location. Mathematical Research Center Technical Report #2102, University of Wisconsin-Madison.

- MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51-67.
- WIENS, D. P. (1982). Robust estimation of multivariate location and scale in the presence of asymmetry. Ph.D. thesis, University of Calgary.
- WIENS, D. P. and ZHENG, Z. (1983). Robust M-estimators of multivariate location and scatter in the presence of asymmetry. Research Report No. DALTR-84-5, Dalhousie University, Halifax, Nova Scotia.

DEPARTMENT OF MATHEMATICS AND STATISTICS  
UNIVERSITY OF CALGARY  
CALGARY, ALBERTA  
CANADA, T2N 1N4

DEPARTMENT OF MATHEMATICS,  
STATISTICS AND COMPUTING SCIENCE  
DALHOUSIE UNIVERSITY  
HALIFAX, NOVA SCOTIA  
CANADA B3H 4H8