MINIMIZING FISHER INFORMATION OVER MIXTURES OF DISTRIBUTIONS

BY

P. J. BICKEL AND J. R. COLLINS

TECHNICAL REPORT NO. 8
MARCH 1982

RESEARCH PARTIALLY SUPPORTED
BY
OFFICE OF NAVAL RESEARCH CONTRACT N0001475C0444
AND
NATURAL SCIENCES AND ENGINEERING RESEARCH COUNCIL
OF CANADA GRANT A-4499

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

MINIMIZING FISHER INFORMATION OVER MIXTURES OF DISTRIBUTIONS

by

P.J. Bickel[1,2] & J.R. Collins[3]

University of California, Berkeley, and University of Calgary

SUMMARY: We study the problem of minimizing Fisher information over the class of distributions of the form $\int F(\cdot, \gamma)\, \nu(d\gamma)$, where $F(\cdot, \gamma)$, $-\infty \leq \gamma \leq \infty$, is a parametric family of distributions and ν is an arbitrary probability measure on $E \subset [-\infty,\infty]$. We show, under suitable regularity conditions on $F(\cdot, \gamma)$, that the ν minimizing Fisher information concentrates all its mass on a countable subset of isolated points in $(-\infty, \infty)$ plus possibly $\{\pm \infty\}$. Applications are given to problems arising in robust estimation, robust smoothing, and minimax MSE estimation of the mean of a normal distribution under various constraints.

[1] Part of this work was performed while the author was enjoying the hospitality and support of the Statistics Department of the Hebrew University of Jerusalem.

[2] Research partly supported by contract N0001475C0444 of the Office of Naval Researcl

[3] Research supported by the Natural Sciences and Engineering Research Council of Canada under Grant No. A-4499. Part of this work was performed while the author was enjoying the hospitality of the Department of Statistics, University of California, Berkeley.

MINIMIZING FISHER INFORMATION OVER MIXTURES OF DISTRIBUTIONS

By P. J. Bickel[1,2] and J. R. Collins[3]

1. INTRODUCTION

Huber (1964) presented a general theory for minimizing the Fisher information for location $I(F)$ where F ranges over a convex and vaguely compact set of distribution functions $\mathcal{F}$ on $\bar{R} = [-\infty,\infty]$. Here

$$(1.1) \qquad I(F) = \begin{cases} \int[(f')^2/f]dx & \text{if the restriction of } F \text{ to } R \text{ has an absolutely continuous density } f, \\ \infty & \text{otherwise}. \end{cases}$$

In this paper, we apply Huber's theory to classes of distributions of the form

$$(1.2) \qquad \mathcal{F} = \left\{ \int_E F(\cdot,\gamma)\nu(d\gamma) : \nu \text{ is an arbitrary probability measure on } \bar{R} \right\},$$

where $E = [a,b]$, $-\infty \leq a < b \leq \infty$; $F(\cdot,\gamma)$, $-\infty < \gamma < \infty$, is a suitably regular parametric family of distribution functions, (and $F(\cdot,\pm\infty)$ are defined by continuity). $E = \bar{R}$ is most interesting.

In Section 2 we show (under suitable regularity conditions on the family $F(\cdot,\gamma)$), that the probability measure ν_0 at which the Fisher information is minimized is concentrated on a countable subset of E. Furthermore, the only possible accumulation points of the support of ν_0 are $\pm\infty$.

In Section 3 we apply these results to several examples, including the following:

(i) $\quad E = \bar{R} : F(x,\gamma) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(xe^{-\gamma})$, $\quad -\infty < \gamma < \infty$, where $0 < \varepsilon < 1$ and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$, where $\phi(t) = (2\pi)^{-\frac{1}{2}} \exp(-t^2/2)$;

(ii) $E = \bar{R}$: a) $F(x,\gamma) = (1 - \varepsilon)\Phi(x) + \varepsilon\Phi(x - \gamma)$, $-\infty < \gamma < \infty$

and more generally

$$\text{b)} \quad F(x,\gamma) = (1 - \varepsilon)\Phi * \pi_0(x) + \varepsilon\Phi(x - \gamma)$$

where π_0 is a fixed distribution.

(iii) $F(x) = \Phi(x - \gamma)$.

These examples arise naturally in two distinct contexts, Huber's asymptotic theory of robust estimation of the mean of a normal distribution and constrained minimax estimation of the mean of a normal distribution.

Thus, Huber (1964) showed that if F minimizes $I(F)$ over $\mathcal{F}$, then $\psi_0 = -f_0'/f_0$ solves the problem of finding ψ which minimizes $\sup\{V(\psi, F) : F \in \mathcal{F}\}$, where $V(\psi, F)$ is a functional which, when defined on the sub-class of F 's with absolutely continuous density f , is given by

$$V(\psi, F) = \int \psi^2 f \, dx / (\int \psi f' \, dx)^2 \quad . \tag{1.3}$$

In the robust estimation problem, one observes a random sample $X_1, \ldots, X_n$ from $F(x - \theta)$, where F is an unknown member of a fixed convex and vaguely compact set $\mathcal{F}$ of distribution functions symmetric about 0 , and θ is an unknown parameter to be estimated. Under regularity conditions on ψ and F , M - estimators of θ based on ψ (i.e., suitably-chosen solutions of $\sum_{i=1}^{n} \psi(X_i - \hat{\theta}_n) = 0$) are seen to be consistent estimators of θ , with $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ converging in distribution to the normal distribution with mean 0 and variance $V(\psi, F)$. In this context, the M - estimator based on

$\psi_0 = -f_0'/f_0$ is most robust for the model $\{F(x - \theta) : F \in \mathcal{F}\}$ in the sense that it achieves minimax asymptotic variance.

A motivation for studying Example (i) is as follows. In Andrews et al. (1972), the Monte Carlo study of robust estimates of a location parameter is restricted to error distributions which are scale mixtures of normal distributions centered at 0 , i.e., to error distributions of the form

$$(1.4) \qquad F(x) = \int \Phi(xe^{-\gamma})\, \mu(d\gamma) \ ,$$

where μ is a probability measure on $[-\infty, \infty]$. This is a broad class of symmetric unimodal distributions; examples are the t , double exponential and logistic distributions [Andrews and Mallows (1974)]. Efron and Olshen (1978) ask the question "How broad is the class of normal scale mixtures?" and answer it by computing the extrema of $F(x_3)$ over distributions of form (1.4) satisfying $F(x_1) = \alpha_1$ and $F(x_2) = \alpha_2$. Another possible approach to the question is to consider Huber's minimax asymptotic variance problem when $\mathcal{F}$ consists of all F of the form (1.4). However this class of distributions is so large that the problem breaks down in the sense that $\inf\{I(F) : F \in \mathcal{F}\} = 0$. If the model is modified so that μ is allowed to vary only over the convex sub-class of probability measures of the form $\mu = (1 - \varepsilon)\delta_1 + \varepsilon\nu$, where $0 < \varepsilon < 1$ and δ_1 denotes unit mass at 1 , then one obtains a class of distributions (Example (i)) for which Huber's minimax variance problem is meaningful:

$$(1.5) \qquad \mathcal{F}_1 = \{F : F(x) = (1 - \varepsilon)\Phi(x) + \varepsilon\int\Phi(xe^{-\gamma})\ \nu(d\gamma) : \nu \text{ is a probability measure on } \bar{R}\}$$

Note that $\mathcal{F}_1$ is contained in the class

$$(1.6) \qquad \mathcal{F}^* = \{F : F = (1 - \varepsilon)\Phi + \varepsilon G \text{ , for some (possibly sub-stochastic) distribution } G \text{ symmetric about } 0\} .$$

for which the minimax variance problem was solved by Huber (1964). Since (i) I(F) is minimized over $\mathcal{F}^*$ at a unique distribution $F_0^* = (1 - \varepsilon)\Phi + \varepsilon G_0$, where $G_0\{[-k, k]\} = 0$ for some $k > 0$, and (ii) $\int \Phi(xe^{-\gamma})\nu(d\gamma)$ places positive mass on $[-k, k]$ except for the special substochastic case where $\nu\{+\infty\} = 1$, it follows that $\inf\{I(F) : F \in \mathcal{F}_1\} > I(F_0^*)$. In particular, the minimax $\psi_0 = -f_0'/f_0$ (corresponding to the F_0 in $\mathcal{F}_1$ minimizing I(F) over $\mathcal{F}$) cannot coincide with Huber's solution $\psi(x) = x$ for $|x| \leq k$, $= k \operatorname{sgn}(x)$ for $|x| > k$. One way to measure how "broad" is the class of normal scale mixtures would be to compute (for various ε) how close $\inf\{I(F) : F \in \mathcal{F}_1\}$ is to $\inf\{I(F) : F \in \mathcal{F}^*\}$. Although the problem of computing the least favorable F_0 remains open we establish the following qualitative results in Section 3: the support of the least favorable ν_0 is a denumerable set with an accumulation point at $+\infty$. Also $\nu_0(+\infty) = 0$, so that F_0 is a proper distribution, i.e., $\int_{-\infty}^{\infty} f_0(x)\,dx = 1$.

Example (ii)a), the problem of minimizing I(F) over

$$(1.7) \qquad \mathcal{F}_2 = \{F: F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\int\Phi(x-\gamma)\nu(d\gamma)\ , \text{ where } \nu \text{ is a probability measure on } \bar{R}\}$$

arises as a problem in the asymptotic theory of robust non-linear smoothing of time series studied by C. Mallows -- for details, see Mallows (1978) and Mallows (1980) p.711.

Bickel (1980), (1981) and Levit (1979, 1980) and Marazzi (1980) independently found an interesting connection between the problems of constrained minimax estimation with quadratic loss of the mean θ of a normal distribution with known variance and minimization of Fisher information. Specifically, if without loss of generality we take sample size 1 and variance 1, the Bayes risk for any prior Π is given by,

$$(1.8) \qquad r(\pi) = 1 - I(\Phi * \pi)$$

where $*$ denotes convolution. Moreover, the Bayes estimate δ_π is well known to be given by

$$(1.9) \qquad \delta_\pi(x) = x + \frac{f_\pi'}{f_\pi}(x)$$

where f_π is the density of $\Phi * \pi$.

The constrained minmax problem of Bickel (1980) is to find an estimate such that $\sup_\theta E_\theta(\delta - \theta)^2$ is minimized subject to $E_0(\delta^2) \le 1 - t$, $t > 0$. He showed this was equivalent to minimizing $\sup_\theta\{(1 - \varepsilon)E_0\delta^2 + \varepsilon \ E_\theta(\delta - \theta)^2\}$ for some $\varepsilon(t)$. By standard minmax arguments this is equivalent to finding $\tilde{\pi}$ and the corresponding $f_{\tilde{\pi}}$ maximizing $1 - I(\Phi * \pi)$ for $\pi = (1 - \varepsilon)\delta_{\{0\}} + \varepsilon\,\tilde{\gamma}$ where γ is arbitrary. This is just example (ii)a). Similarly Efron and Morris (1971) and Marazzi considered the problem of minimizing $\sup_\theta E_\theta(\delta - \theta)^2$ subject to $\int E_\theta(\delta - \theta)\pi_0(d\theta) \le c$ for π_0 a fixed Bayes prior distribution. This leads to example (ii) b).

Finally, consider the problem of minimizing $\sup\{E_\theta(\delta - \theta)^2 : \theta \in E\}$ studied by various workers. Correspondence (1.8) and a standard arguement shows that this problem is equivalent to minimizing $I(F)$ as in example (iii) Some results on this problem can be found in Casella and Strawderman (1981), and Bickel (1981).

2. THE GENERAL RESULT

Let $\mathcal{F}$ denote a class of distribution functions

$$(2.1)\qquad \mathcal{F} = \{ \int_{\bar{E}} F(\cdot, \gamma)\nu(d\gamma) : \nu \text{ ranges over all probability measures on } \bar{R}\},$$

where $E = [a, b] \subset \bar{R} = [-\infty,\infty]$, and $F(\cdot, \gamma)$ is a fixed parametric family of distribution functions satisfying the following conditions:

(A.1) $F(\cdot, \gamma)$ is absolutely continuous with density $f(\cdot, \gamma)$ for all $\gamma \in R$;

(A.2) For every compact set $K \subset R$, there is a number $c_K > 0$ such that $f(x, \gamma) \geq c_K$ for all $x \in K$ and for all $\gamma \in R$;

(A.3) For all $\gamma \in R$, $f(\cdot,\gamma)$ is absolutely continuous with density $f'(\cdot, \gamma)$; such that

(A.4) Both $\int |f'(x, \gamma)| dx$ and $\int |xf'(x,\gamma)| dx$ are finite for all $\gamma \in R$.

Let f_+ and f_- be defined by

$$f_+(x, \gamma) = - \int_x^\infty [f']^+(y, \gamma)dy$$

$$f_-(x, \gamma) = - \int_x^\infty [f']^-(y, \gamma)dy \quad .$$

It is easy to see that (A.3) and (A.4) imply,

$$f = f_+ - f_-$$

(A.5) $\int |w(x)| f(x, \gamma)dx < \infty$ if and only if both $\int | w(x)| f_+(x, \gamma)dx < \infty$ and $\int |w(x)| f_-(x, \gamma)dx < \infty$, where f_+ and f_- are as defined above.

(A.6) $I(F(\cdot,\gamma)) < \infty$ for all $\gamma \in R$; and

(A.7) $\inf\{I(F) : F \in \mathcal{F}\} > 0$.

(B.1) The functions $f(x, \gamma)$ and $f'(x, \gamma)$ are analytic in γ (a.e.x) ;

(B.2) The functions $\int w(x)f(x, \gamma)dx$ and $\int w(x)f'(x, \gamma)dx$ are analytic in γ on any interval of γ 's for which they are well-defined and finite;

(B.3) For $i = 0$ or 1 , $\int |w(x)||f^{(i)}(x, \gamma)|dx < \infty$ for $\gamma = a$ and $\gamma = b$ implies that $\int |w(x)||f^{(i)}(x, \gamma)|dx < \infty$ for all $\gamma \in [a, b]$.

(C.1) $F(\cdot, \gamma)$ converges in distribution to $F(\cdot, \pm \infty)$ as $\gamma \to \pm \infty$;

(C.2) The functions $f(x, \pm \infty)$ defined by $f(x, \pm \infty) = \lim_{\gamma \to \pm\infty} f(x, \gamma)$ exist (a.e.x) . If the restriction of $F(\cdot, +\infty)$ resp. $F(\cdot, -\infty)$ to R is absolutely continuous then $f(\cdot, +\infty)$ resp. $f(\cdot, -\infty)$ is its density and is itself absolutely continuous with derivative which we write $f'(\cdot, \pm \infty)$.

Note: Condition C is unnecessary unless either $a = -\infty$ or $b = \infty$.

(D) Either

(i) $\int w(x)\ f'(x, \gamma)\ dx = 0$ for all $c < \gamma < d$, implies w is constant a.e.

or

(ii) $f(x, \gamma)$ symmetric in x , $\int w(x)f'(x, \gamma)dx = 0$ for all $c < \gamma < d$, w antisymmetric in x imply $w(x) = 0$ a.e.x.

(E) $\frac{d}{dx} \int_E f(x, \gamma)\nu(d\gamma) = \int_E f'(x, \gamma)\nu(d\gamma)$ whenever either side exists and is finite.

(F) Either,

(i) $\int |f'(x, \gamma)|\nu(d\gamma) \leq c(\nu)|\int_E f'(x, \gamma)\nu(d\gamma)|$ for $c(\nu) < \infty$, all x and all ν, or

(ii) $\sup_E I(F(\cdot, \gamma)) < \infty$.

Define $J(\psi, \gamma)$ for $\gamma \in E$ by:

(2.2) $$J(\psi, \gamma) = \int[-2\psi(x)f'(x, \gamma) - \psi^2(x) f(x, \gamma)]\, dx$$

if $$\int\psi^2(x) f(x, \gamma)\, dx < \infty$$

$$= \infty \quad \text{otherwise}$$

Note that the integral on the right in (2.2) exists and is finite $\Longleftrightarrow \int \psi^2(x) f(x, \gamma)\, dx < \infty$ since $I(F(\cdot, \gamma)) < \infty$, $\forall \gamma$. To see the implication, let ψ_M be the truncation of ψ at $\pm M$ and note that

$$J(\psi, \gamma) = \lim_M \int[2\psi_M(x) f'(x, \gamma)\, dx - \psi_M^2(x) f(x, \gamma)]\, dx$$

$$\leq \lim_M([\int\psi_M^2(x) f(x, \gamma)\, dx]^{\frac{1}{2}} (2I^{\frac{1}{2}}(F(\cdot, \gamma)))$$

$$- [\int \psi_M^2(x) f(x,\gamma)]dx) \to -\infty$$

if $\int\psi^2(x) f(x, \gamma)\, dx = \infty$. The converse is immediate.

Theorem 1: Under condition (A) - (F), $I(F)$ is minimized over $\mathcal{F}$ by a probability distribution $F_0 = \int_E F(\cdot, \gamma)\nu_0(d\gamma)$, such that

(i) F_0 is unique

(ii) The density f_0 of F is absolutely continuous and $\psi_0 = -\dfrac{f_0'}{f_0}$ exists a.e.

(iii) The support of ν_0 is contained in $\{\gamma : J(\psi_0, \gamma) = I(F)\}$

(iv) The support of ν_0 is a countable subset of E with $\pm\infty$ as the only possible accumulation points.

Before giving the proof of theorem 1 we briefly review Huber's (1964,1977) results.

Fisher information $I(F)$ is a convex functional attaining its minimum over $\mathcal{F}$ at an F_0 with absolutely continuous density f_0. A necessary and sufficient condition for F_0 to minimize $I(F)$ over $\mathcal{F}$ is that $\frac{d}{d\varepsilon} I[(1-\varepsilon)F_0 + \varepsilon F_1]\Big|_{\varepsilon=0} \geq 0$ for all $F_1 \in \mathcal{F}$, or equivalently that

$$\int [-2\psi_0 f_1' - \psi_0^2 f_1]\, dx \geq I(F_0) \tag{2.3}$$

for all $F_1 \in \mathcal{F}$ which have absolutely continuous density f_1, where the function ψ_0 is defined by $\psi_0 = -f_0'/f_0$. Our study consists of ascertaining the implications of Huber's condition (2.3) when $\mathcal{F}$ is of the form (1.2).

Proof: Evidently $\mathcal{F}$ is convex and also weakly compact since $F(x, \gamma)$ is bounded and continuous in γ (by (B.2) and (C.1)). Since, by (A.6), $I(F) < \infty$ for some $F \in \mathcal{F}$ the existence of F_0 is guaranteed by Huber. Its unicity is guaranteed by (A.2), (Huber (1977)) and (i) and (ii) are established.

To establish (iii) note first that by (A.1), (A.3), (C.1), (C.2) and (E) $F(\cdot, \nu) \in \mathcal{F}$ and $I(F(\cdot, \gamma)) < \infty$ imply that the restriction of $F(\cdot, \nu)$ to R has an absolutely continuous density $f(\cdot, \nu)$ with Radon-Nikodym derivative $f'(\cdot, \nu)$ given by,

$$f(x, \nu) = \int_E f(x, \gamma)\nu(d\gamma) \tag{2.4}$$

$$f'(x, \nu) = \int_E f'(x, \gamma)\nu(d\gamma) . \tag{2.5}$$

Thus, (2.3) implies that,

$$\int [-2\psi_0(x) f'(x, \nu) - \psi_0^2(x) f(x, \nu)]\, dx \geq I(F_0) \tag{2.6}$$

for all ν such that $I(F(\cdot, \nu)) < \infty$. In particular for J defined by (2.2),

$$J(\psi_0, \gamma) \geq I(F_0) \tag{2.7}$$

for all $\gamma \in E$.

We now show that,

(2.8) $$\int_E J(\psi_0, \gamma)\nu_0(d\gamma) = I(F_0)$$

which will establish (iii).

Note that the left-hand side of (2.8) is well-defined because $J(\psi_0, \gamma)$ is bounded below. Clearly it suffices to show that

(2.9) $$\int[\int \psi_0^2(x)f(x,\gamma)dx]\nu_0(d\gamma) = \int\psi_0^2(x)[\int f(x,\gamma)\nu_0(d\gamma)]\, dx$$

(2.10) $$\int[\int\psi_0(x)f'(x,\gamma)dx]\nu_0(d\gamma) = \int\psi_0(x)[\int f'(x,\gamma)\nu_0(d\gamma)]\, dx$$

since then

$$\int J(\psi_0,\gamma)\nu_0(d\gamma) = \int(2\psi_0^2 - \psi_0^2)f(x,\nu_0)dx = I(F_0) \quad .$$

Since the right-hand side of (2.9) is $I(F_0) < \infty$, and $\psi_0^2(x)f(x, \gamma)$ is non-negative, (2.9) is immediate. To prove (2.10) we need only show that

(2.11) $$\iint|\psi_0(x)||f'(x, \gamma)|\, dx\, \nu_0(d\gamma) < \infty \quad .$$

We claim that (2.11) follows from either condition (F) (i) or F (ii).

First suppose that condition (F)(i) holds. Then

(2.12) $$\infty > I(F_0) = \int|\psi_0(x)||\int f'(x, \gamma)\nu_0(d\gamma)|\, dx$$

$$\geq \inf_x \frac{|\int f'(x, \gamma)\nu_0(d\gamma)|}{\int|f'(x, \gamma)|\nu_0(d\gamma)} \cdot \int|\psi_0(x)|[\int|f'(x,\gamma)|\nu_0(d\gamma)]dx$$

$$= \frac{1}{c(\nu_0)}\iint|\psi_0(x)||f'(x, \gamma)|\nu_0(d\gamma)\, dx ,$$

so that (2.11) holds. Now suppose instead that condition F(ii) holds. Then, by the Cauchy-Schwartz inequality,

$$(2.13)\qquad \iint |\psi_0(x)|\,|f'(x,\gamma)|\, dx\ \nu_0(d\gamma) = \iint |\psi_0(x)| \frac{|f'(x,\gamma)|}{f(x,\gamma)} f(x,\gamma)dx\ \nu_0(d\gamma)$$

$$\leq [\iint \psi_0^2(x)f(x,\gamma)dx\ \nu_0(d\gamma)]^{\frac{1}{2}} \cdot [\iint \frac{|f'(x,\gamma)|^2}{f(x,\gamma)}\ dx\ \nu_0(d\gamma)]^{\frac{1}{2}}$$

$$\leq [I(F_0)]^{\frac{1}{2}} [\sup_{\gamma} I(F(\cdot,\gamma))]^{\frac{1}{2}} < \infty .$$

This completes the verification of (2.8) and hence (iii).

We now prove (iv). Note that $\infty > I(F_0) = \iint \psi_0^2(x)f(x,\gamma)dx\ \nu_0(d\gamma)$ implies, by condition (B.3), that $\int \psi_0^2(x)f(x,\gamma)dx < \infty$ for all γ in K_{ν_0}, the convex hull of the support of ν_0. By (B.2), $\int \psi_0^2(x)f(x,\gamma)dx$ is analytic in γ on K_{ν_0}, a.e.x. Also, by Cauchy-Schwartz, $\int |\psi_0(x)|\,|f'(x, \gamma)|\ dx \leq [\int \psi_0^2(x)f(x,\gamma)dx]^{\frac{1}{2}} [I(F(\cdot,\gamma))]^{\frac{1}{2}}$, so that by (A.6) $I(F(\cdot,\gamma) < \infty$, (B.2) and (B.3), it follows that $\int \psi_0(x)f'(x,\gamma)dx$ is analytic in γ on K_{ν_0}, a.e.x. Thus $J(\psi_0, \gamma)$ is (a.e.x) an analytic function on K_{ν_0}, and the support of ν_0 is contained in a subset of $\bar{R}$ on which an analytic function attains its minimum.. There are now two possibilities:

(a) (iv) holds; or

(b) $J(\psi_0, \gamma) \equiv I(F_0)$ for all $\gamma \in K$.

Otherwise, the support of ν_0 must have an accumulation point in $(-\infty, \infty)$. But since the analytic function $J(\psi_0, \gamma)$ then is constant on a set containing a limit point in $(-\infty, \infty)$, $J(\psi_0, \gamma)$ must necessarily be constant everywhere on K_{ν_0}, so that (b) holds.

We claim that (b) is impossible, and hence, establish (iv). We begin by showing that

$$(2.14)\qquad \infty > \int \psi_0^2(x)\ f(x,\gamma)\ dx = -\int W_0(x)f'(x, \gamma)\ dx$$

a.e. ν_0 , where $W_0(x)$ is defined for all x by

$$W_0(x) = \int_0^x \psi_0^2(y)\, dy . \tag{2.15}$$

The inequality has already been shown to hold a.e. ν_0 . To verify the integration-by-parts, first note that $W_0(x) < \infty$ for all x . For if $\int_0^x \psi_0^2(y)dy = \infty$ for some x , then $f_0(y) \geq C > 0$ for all $y \in [0, x]$ by (A.2), so that $\infty > \int \psi_0^2 f_0 \geq C\int_0^x \psi_0^2\,(y)\, dy$, a contradiction. Let $[W_0]^+(x)$ be the positive part of W_0 . Using the notation of condition (A.5) , we have (a.e. ν_0)

$$\int_{-\infty}^{\infty} [W_0]^+(x)[f']^+(x, \gamma)\, dx = \int_0^{\infty}[\int_0^x \psi_0^2(y)\, dy][f']^+(x, \gamma)\, dx \tag{2.16}$$

$$= \int_0^{\infty}\psi_0^2(y)[\int_y^{\infty}[f']^+(x, \gamma)\, dx]\, dy = -\int_0^{\infty}\psi^2(x)f_+(x, \gamma)\, dx .$$

It follows from similar calculations using integrands $[W_0]^-[f']^+$, $[W_0]^+[f']^-$ and $[W_0]^-[f']^-$, and from conditions (A.5) and (A.6) , that (2.14) holds. Thus condition (b) is equivalent to

$$\int[-2\psi_0(x) + W_0(x)]\, f'(x, \gamma)\, dx = I(F_0) \tag{2.17}$$

for all $\gamma \in K_{\nu_0}$. In view of (A.4), integration by parts gives,

$$\int xf'(x, \gamma)\, dx = -1 , \tag{2.18}$$

Thus (2.17) and (2.18) give,

$$\int[-2\psi_0(x) + W_0(x) + [I(F_0)]x]\, f'(x, \gamma)\, dx = 0 \tag{2.19}$$

for all $\gamma \in K_{\nu_0}$. If condition (D) (i) holds, then

$$Q(x) = -2\psi_0(x) + \int_0^x \psi_0^2(y)\, dy + [I(F_0)]\, x = C \text{ a.e.} \tag{2.20}$$

If instead condition D(ii) holds, then $Q(x) = Q(-x)$ a.e. since $f(x, \gamma)$

is symmetric for all ν. So $Q(x) = 0$ a.e. i.e., (2.20) is also satisfied, with $C = 0$ under condition (D) (ii). It follows from (2.20) that ψ is differentiable and satisfies

$$(2.21) \qquad 2\,\psi_0'(x) - \psi_0^2(x) = +I(F_0) > 0\,. \quad \text{a.e.x.}$$

The only solutions to (2.21) are of the form $\psi(x) = \lambda \tan \frac{1}{2}\lambda(x-c)$ where $\lambda = I(F_0)^{\frac{1}{2}}$, valid on the interval $C \pm \frac{2\pi}{\lambda}$. So by (A.2) there exists no absolutely continuous density function f_0 for which $\psi_0 = -f_0'/f_0$ satisfies (2.21) a.e.x. We have a contradiction and (iv) follows. □

<u>Remark 2.1</u> The support of ν_0 in the conclusion of Theorem 1 may be either a finite or a denumberable set of points. Noting that, under the conditions of Theorem 1,

$$(2.22) \qquad \sup_{F \in \mathcal{F}} V(\psi_0, F) = 1/[I(F_0)] < \infty\,,$$

(where $V(\psi,F)$ is defined by (1.3)), we see that a sufficient condition for the support of ν_0 to be a denumberable set is as follows: for every $f^* = \int f(x, \gamma)\, \nu^*(d\gamma)$ for which ν^* concentrates on a finite set of points, we have

$$(2.23) \qquad \lim_{\gamma} \sup V(\psi^*, F(\cdot, \gamma)) = \infty\,,$$

where $\psi^* = -f^{*\prime}/f^*$. This condition can be easily verified in the two examples in the next section.

3. EXAMPLES

(i) $F(x,\gamma) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(xe^{-\gamma})$, $\gamma \in R$.

Theorem 2: The conditions and conclusions of Theorem 1 are satisfied for example (i). Moreover,

a) ν_0 is a proper probability on R, viz $\nu_0\{-\infty\} = \nu_0\{+\infty\} = 0$.

b) The support of ν_0 is denumerable with an accumulation point at $+\infty$.

Before proving this theorem, we will discuss some of its aspects.

Remark 3.1. The extremal F_0 of Theorem 2 has density of the form

$$f_0(x) = (1-\varepsilon)\phi(x) + \varepsilon\sum_{i=1}^{\infty} p_i\sigma_i^{-1}\phi(x\sigma_i^{-1}) \quad \text{for} \quad 0 < p_i < 1 \quad , \tag{3.1}$$

$\Sigma p_i = 1$ and $0 < \sigma_i < \infty$ for $i = 1,2,\ldots$. Side conditions determining the unique values of the p_i and σ_i can be derived by writing, for a general F of the form (3.1),

$$I(F) = 2\int_0^{\infty} x^2 \frac{\{(1-\varepsilon)\phi(x) + \varepsilon\sum_{i=1}^{\infty} p_i\sigma_i^{-3}\,\phi(x\sigma_i^{-1})\}^2\,dx}{\{(1-\varepsilon)\phi(x) + \varepsilon\sum_{i=1}^{\infty} p_i\sigma_i^{-1}\,\phi(x\sigma_i^{-1})\}} \tag{3.2}$$

and minimizing over σ_i, p_i subject to the side conditions. Through a standard argument with a Lagrange multiplier we arrive at,

$$2\int_0^{\infty} [2x\psi_0(x)\sigma_i^{-2} - \psi_0^2(x)]\sigma_i^{-1}\,\phi(x\sigma_i^{-1})dx = m \quad , \quad i = 1,2,\ldots \tag{3.3}$$

where

$$m = \varepsilon^{-1}[I(F_0) - (1-\varepsilon)\int[2x\psi_0(x) - \psi_0^2(x)]\phi(x)dx] < 0$$

and

$$(3.4) \qquad \int_0^\infty [2x\psi_0(x)\sigma_i^{-2}(x^2\sigma_i^{-2} - 3) - \psi_0^2(x)(x^2\sigma_i^{-2} - 1)]\ \phi(x\sigma_i^{-1})dx = 0 \quad , \quad i = 1,2,\ldots .$$

We do not know explicit formulae for the (p_i,σ_i) nor even whether the optimal (p_i,σ_i) are the unique solutions of this system of equations. Nor are we suggesting that this is the approach to numerical solution of our problem. What may be more appropriate are algorithms of the type introduced in the theory of convex programming -- see for example Wu (1978).

Remark 3.2. The optimal ψ_0 is clearly antisymmetric and satisfies $\psi_0(x) > 0$ for all $x > 0$. It is also shown in the proof of Theorem 2 that $\psi_0(x)/x$ is monotone non-increasing for $x > 0$. However, it is not known whether ψ_0 is bounded or even whether ψ_0 is continuous at $x = 0$. (For note that Theorem 2 leaves open the question of whether $-\infty$ is an accumulation point of the support of ν_0.)

Remark 3.3. The conditions for Theorem 1 are undoubtedly too strong. However, the conclusion can fail without some smoothness conditions on $F(\cdot\,,\cdot)$. A striking example is provided by the unimodal contamination model considered by Donoho (1978), i.e., $\mathcal{F} = \{F : F = (1-\varepsilon)\Phi + \varepsilon G,\ G$ symmetric unimodal$\}$ which may also be written as a family of the form (2.1) with $F(x,\gamma) = (1-\varepsilon)\Phi(x) + \varepsilon\, U(xe^{-\gamma})$ where U is the uniform distribution on $[-1,1]$. To see that the conclusion of Theorem 1 must fail here, note that if ν is any probability measure on $\bar{R}$ for which the support is a denumerable set with only $-\infty$ and ∞ as possible limit points, then the corresponding $F = \int F(x, \gamma)\ \nu(d\gamma)$ has a density which is not absolutely continuous, so that $I(F) = \infty$. In the solution to this particular problem, the least favorable ν_0 turns out to have an absolutely continuous component. For the solution, see Donoho (1978). Of course (A) - (F) are violated in various ways. Since $F(x, \gamma)$ is not continuous in x , $I(F(\cdot, \gamma) = \infty$ for all γ, and $f(x, \gamma)$ is not a continuous function of γ , much less analytic.

Remark 3.4: If E is a bounded interval and theorem 1 holds the carrier of ν_0 is necessarily finite. It is trivially clear that we can reparametrize $F(\cdot,\gamma)$ so that ν_0 is carried by a finite set even though γ ranges over $\bar{R}$. For instance, define

$$f(\cdot, \gamma) = (1 - \varepsilon)\ \phi(x) + \varepsilon\ \lambda(\gamma)\ \phi(x\ \lambda(\gamma)\ )$$

where λ is an analytic $1 - 1$ mapping of $\bar{R}$ onto (say) $[0,1]$. For instance we can take $\lambda = \Phi$.

Proof of theorem 2: The verification of conditions (A.1) - (A.5) is elementary. Assumption (A.6) follows since I is a convex functional, so that,

$$I(F(\cdot, \gamma)) \leq (1 - \varepsilon)\ I(\Phi) + \varepsilon\ I(\Phi(xe^{-\gamma}))$$

$$= (1 - \varepsilon) + \varepsilon\ e^{-2\gamma} < \infty .$$

The argument for (A.7) was given in section 1.

Conditions (B.1) - (B.3) follow from standard properties of the Laplace transform once we represent

$$\int |w(x)|\ \exp(-\frac{x^2}{2\sigma^2})\ dx = \tfrac{1}{2} \int_0^\infty \left(|\ w(y^{\frac{1}{2}})| + |w(-y^{\frac{1}{2}})|\right)\ y^{-\frac{1}{2}}\ \exp\left(-\frac{y}{2\sigma^2}\right) dy .$$

Conditions (C.1) and (C.2) hold:

Here,

$$F(\cdot, -\infty) = (1-\varepsilon)\Phi + \varepsilon\, \delta_{\{0\}}$$

$$F(\cdot, +\infty) = (1-\varepsilon)\Phi + \varepsilon(\delta_{\{\infty\}} + \delta_{\{-\infty\}}) \quad .$$

In this case since the restriction of $F(\cdot, -\infty)$ to R is not absolutely continuous $I(F(\cdot, -\infty)) = \infty$. On the other hand, the restriction of $F(\cdot, +\infty)$ is absolutely continuous, has density $(1-\varepsilon)\phi(x) = f(x, \infty)$, with $I(F(\cdot, \infty)) = (1-\varepsilon)$.

To verify condition D(ii) note that $f(x, \gamma)$ is symmetric in x . If $\int w(x)\, f'(x, \gamma)dx = 0$, $\gamma \in (a, b]$ both $\int |x\, w(x)| e^{-\frac{1}{2}x^2}\, dx$ and $\int |x\, w(x)| e^{-3\gamma} \exp(-\frac{1}{2} x^2 e^{-2\gamma})\, dx$ converge and thus $\int xw(x) e^{-3\gamma} \exp(-\frac{1}{2}x^2 e^{-2\gamma})\, dx$ is finite and analytic in γ for a region S containing (a, b) and 0 . Therefore,

$$\text{(3.5)} \qquad \int xw(x)\, e^{-3\gamma} \exp(-\frac{x^2}{2} e^{-2\gamma})dx = -\frac{(1-\varepsilon)}{\varepsilon} \int xw(x) e^{-\frac{x^2}{2}}\, dx$$

for all $\gamma \in S$. Putting $\gamma = 0$ shows that the right hand side of (3.5) is 0 . A change of variables leads to

$$\text{(3.6)} \qquad \int_0^\infty [w(\sqrt{y}) - w(-\sqrt{y})] \exp(-y\, e^{-2\gamma})\, dy = 0$$

for $\gamma \in S$. The unicity theorem for Laplace transforms completes the verification of D(ii).

(E) follows similarly from the classical formula for differentiation of Laplace transforms.

Condition (F)(i) is easily seen to hold with $C(\nu) = 1$ for all ν.

To complete the proof of the theorem we need to verify a) and b).

At this point it is convenient to re-parametrize by defining $\sigma = e^{\gamma}$. Write $G(x, \sigma)$ for $F(x, \gamma)$ and $G(x, \mu) = \int_0^{\infty} G(x, \sigma)\, \mu(d\sigma)$ for $F(x, \nu)$ with the correspondence $\mu(d\sigma) = \nu(d(\log \sigma))$. Also let $\mu_0(d\sigma) = \nu_0(d(\log \sigma))$.

Proof of a): First note that $\mu_0\{0\} = 0$ since $I(F(\cdot, -\infty)) = \infty$.

Now suppose that $\mu_0\{+\infty\} > 0$. Then there is a number $\alpha > 0$ such that *every* density of form

$$f_1(x) = f_0(x) + \alpha\varepsilon \int \frac{1}{\sigma}\, \phi\left(\frac{x}{\sigma}\right) \mu(d\sigma) \tag{3.7}$$

is the density of an $F_1 \in \mathcal{F}$. Since $\alpha\varepsilon > 0$, Huber's condition (2.3) $\int [-2\psi_0(f_1' - f_0') - \psi_0^2(f_1 - f_0)] \geq 0$, when applied to all f_1 of form (3.7) yields

$$\int [-2\psi_0 g' - \psi_0^2 g]\, dx \geq 0 \tag{3.8}$$

$$\text{for all } g = G', \quad G \text{ of the form } \int \Phi\left(\frac{x}{\sigma}\right) \mu(d\sigma) .$$

In particular, for each $\lambda > 0$, there is a G of this form with $g(x) = G'(x) = \frac{1}{2}\lambda \exp(-\lambda|x|)$ [see Andrews and Mallows (1974)]. Also, since ψ_0 is antisymmetric and $\psi_0(x) \operatorname{sgn}(x) \geqslant 0$, it follows from (3.8) that

$$\int_0^{\infty} [2\psi_0(x)\lambda^2 \exp(-\lambda x) - \psi_0^2(x)\lambda \exp(-\lambda x)]dx \geqslant 0 \quad \text{for all } \lambda > 0. \tag{3.9}$$

Using the notation $E_\lambda(h) = \int_0^{\infty} h(x)\, \lambda \exp(-\lambda x)dx$, (3.9) may be written as

$$E_\lambda[2\lambda\psi_0 - \psi_0^2] \geqslant 0 \qquad \text{for all } \lambda > 0 . \tag{3.10}$$

Since $E_\lambda\psi_0 > 0$, (3.10) can be written as $E_\lambda\psi_0^2/[E_\lambda\psi_0] \leqslant 2\lambda$ for all $\lambda > 0$.

Since $0 < \mathrm{Var}_\lambda(\psi_0) = E_\lambda\psi_0^2 - [E_\lambda\psi_0]^2$, it follows that $0 < [E_\lambda\psi_0^2/(E_\lambda\psi_0)] - E_\lambda\psi_0$, so that

$$0 < E_\lambda\psi_0 < \frac{E_\lambda\psi_0^2}{E_\lambda\psi_0} \leq 2\lambda \qquad \text{and}$$

$$E_\lambda\psi_0^2 \leq 2\lambda E_\lambda\psi_0 \leq 4\lambda^2 \qquad \text{for all } \lambda > 0 \quad .$$

Therefore, for all $\lambda > 0$,

$$\int_0^\infty \psi_0^2(x)\, e^{-\lambda x}\, dx \leq 4\lambda \quad . \tag{3.11}$$

Fix $\lambda_0 > 0$. By (3.11) and monotone convergence, letting $\lambda \downarrow 0$, $\int_0^\infty \psi_0^2(x)dx \leq 4\lambda_0$. Since λ_0 is arbitrary, $\int_0^\infty \psi_0^2(x)dx \leq 0$ leading to $\psi_0(x) = 0$ a.e., a contradiction. It follows that $\mu_0\{+\infty\} = 0$.

<u>Proof of b)</u>: We define

$$w(x) = \frac{\psi_0(x)}{x}$$

and then establish that w satisfies $w(x) \geq 0$ for all $x > 0$ and is monotone non-increasing for $x > 0$. Following Andrews and Mallows (1974), a change of variables $\theta = -1/(2\sigma^2)$ yields that

$$f_0(x) = b(x^2)$$

where

$$b(y) = \int_{-\infty}^0 e^{\theta y}\, \nu^*(d\theta) \quad ,$$

where ν^* is a positive measure on $(-\infty,0)$. Since $b(y) < \infty$ for all $y > 0$, it follows that $b(y)$ is analytic for $y > 0$ and that $b^{(n)}(y) = \int_{-\infty}^0 \theta^n\, e^{\theta y}\, \nu^*(d\theta)$ for $y > 0$ and $n = 1,2,\ldots$. Then,

$$\psi_0(x) = [-\log f_0(x)]' = \frac{-2xb'(x^2)}{b(x^2)} \quad ,$$

so that $w(x) \geqslant 0$ for all $x > 0$ since $b'(y) < 0$ for $y > 0$.
To show that $w(x)\downarrow$ as $x \to \infty$, it suffices to show that $[\log\ b(y)]' = b'(y) / b(y)\uparrow$ as $y \to \infty$. But $[\log\ b(y)]'' \geqslant 0$ for all $y > 0$ since

$$(3.12) \qquad [\log\ b(y)]'' = \frac{\int_{-\infty}^{0} \theta^2 e^{\theta y} \nu^*(d\theta)}{\int_{-\infty}^{0} e^{\theta y} \nu^*(d\theta)} - \left[\frac{\int_{-\infty}^{0} \theta e^{\theta y} \nu^*(d\theta)}{\int_{-\infty}^{0} e^{\theta y} \nu^*(d\theta)}\right]^2$$

and the right-hand side of (3.12) is the variance of a random variable. This completes the proof that $w(x)\downarrow$ as $x \to \infty$.

Now suppose that b) is *not* true. Then by (iv) of Theorem 1, the support of ν_0 is either a finite set of points or it is denumerable with a unique accumulation point at $-\infty$. In either case, there is a number σ^*, $0 < \sigma^* < \infty$, such that

$$f_0(x) = (1-\varepsilon)\phi(x) + \varepsilon \sum_i p_i \sigma_i^{-1} \phi(x\sigma_i^{-1}) \quad ,$$

where the index i ranges over a finite or denumerable set, $\sum_i p_i = 1$, $0 < p_i < 1$ for each i, and $0 < \sigma_i \leq \sigma^*$ for each i. Then,

$$(3.13) \qquad w(x) = \frac{(1-\varepsilon)\phi(x) + \varepsilon \sum_i p_i \sigma_i^{-3} \phi(x\sigma_i^{-1})}{(1-\varepsilon)\phi(x) + \varepsilon \sum_i p_i \sigma_i^{-1} \phi(x\sigma_i^{-1})} \quad , \quad x \neq 0 \ ,$$

so that clearly $w(x) \geqslant \min\{1, (\sigma^*)^{-2}\} > 0$ for all $x > 0$. Also since $w(x)\downarrow$, $w(\infty)$ exists and hence $w(\infty) \geqslant \min\{1, (\sigma^*)^{-2}\} > 0$.

For each $\sigma > 0$, let $g(\cdot,\sigma) = (1-\varepsilon)\phi(\cdot) + \varepsilon\sigma^{-1}\phi(\cdot\ \sigma^{-1})$. Then,

$$\int \psi_0^2(x)\ g(x,\sigma)dx = (1-\varepsilon) \int \psi_0^2(x)\phi(x)dx + \varepsilon\sigma^2 \int w^2(\sigma y)y^2\phi(y)dy$$

$$\geqslant (1-\varepsilon) \int \psi_0^2(x)\phi(x)dx + \varepsilon\sigma^2 w^2(\infty) \quad ,$$

so it follows that

(3.14) $$\lim_{\sigma\to\infty} \int \psi_0^2(x)g(x,\sigma)dx = \infty \ .$$

We also have that

(3.15) $$0 < \lim_{\sigma\to\infty} [-\int \psi_0(x)g'(x,\sigma)dx] < \infty \ .$$

To see this, write

$$-\int \psi_0(x)g'(x,\sigma)dx = (1-\varepsilon)\int_{-\infty}^{\infty} x\psi_0(x)\phi(x)dx + 2\varepsilon\int_0^{\infty} w(\sigma y)y^2\phi(y)dy$$

and note that

$$(1-\varepsilon)\int_{-\infty}^{\infty} x\psi_0(x)\phi(x)dx \leqslant -\int_{-\infty}^{\infty} \psi_0(x)f_0'(x)dx = I(F_0) < \infty$$

and that

$$\lim_{\sigma\to\infty} 2\varepsilon\int_0^{\infty} w(\sigma y)y^2\phi(y)dy = \varepsilon w(\infty) < \infty$$

by the dominated convergence theorem, since $w(\sigma y) \downarrow w(\infty)$ as $\sigma \to \infty$ for each $y > 0$, and since for all $\sigma \geqslant 1$, $|w(\sigma y)\, y^2\, \phi(y)|$ is bounded above by $|y\,\psi_0(y)\phi(y)|$.

(ii) $$F(x, \gamma) = (1 - \varepsilon)\Phi(x) + \varepsilon\,\Phi(x - \gamma)$$

We state our result for example (ii)a) for simplicity. However, the proof goes over verbatim to the more general case (ii) b).

Theorem 3: The conditions and conclusions of theorem 1 are satisfied for example (i) . Moreover,

a) ν_0 is symmetric

b) The support of ν_0 is denumerable.

c) ν_0 is a proper probability distribution on R .

Remark 3.5: The solution has the form

$$f_0(x) = (1 - \varepsilon)\phi(x) + \frac{\varepsilon}{2}\sum_{k=1}^{\infty} p_k\,[\phi(x - \gamma_k) + \phi(x + \gamma_k)] ,$$

where $0 < \gamma_1 < \gamma_2 < \ldots < \infty$, $0 < p_k < 1$ for all k and $\sum_{k=1}^{\infty} p_k = 1$. By the argument in Remark 3.1, $\{(p_k, \gamma_k) , i=1,2,\ldots\}$ satisfy the side conditions

$$(3.16) \qquad 2\int_0^{\infty} [2\psi_0'(x) - \psi_0^2(x)]\tfrac{1}{2}[\phi(x - \gamma_k) + \phi(x + \gamma_k)]\,dx = m < 0 \quad \text{for } k=1,2,\ldots$$

and

$$(3.17) \qquad \int_0^{\infty}[2\psi_0'(x) - \psi_0^2(x)][(x - \gamma_k)\,\phi(x - \gamma_k) - (x + \gamma_k)\,\phi(x - \gamma_k)]dx = 0$$

for $k = 1,2,\ldots$

Again, other approaches are appropriate for computation.

Mallows (1980) conjectured that $\gamma_k = hk$ for $k = 1,2,\ldots$ and for some $h > 0$, and that $p_k = \lambda(1 - \lambda)^{k-1}$ for all k for some $\lambda > 0$.

D. Donoho has a modification of this conjecture in which the γ_k 's are of the form $\gamma_k = a + kh$ for some $a > 0$ and $h > 0$. It is not yet known whether either conjecture is true, although numerical results for Donoho's conjecture are encouraging.

Remark 3.6: Another application of this theorem is to the estimation of log scale for contaminated normal families dealt with in Huber (1964) viz.

$$f(x, \gamma) = (1 - \varepsilon) (\tfrac{2}{\pi})^{\frac{1}{2}} \exp \{-\tfrac{1}{2} e^{2y} - y\}$$

$$+ \varepsilon (\tfrac{2}{\pi})^{\frac{1}{2}} \exp\{-\tfrac{1}{2} e^{2(y-\gamma)} - (y - \gamma)\} \quad .$$

Proof: We will verify conditions (D)(i) and (F)(ii) . The verification of the remaining conditions are omitted since they are very similar to the first example (with an additional complication that the set on which $f'(x,\gamma) = [f']^+(x,\gamma)$ depends on γ).

To verify (F)(ii), note that $I(F(\cdot, \gamma)) \leq (1 - \varepsilon)I(\Phi(\cdot)) + \varepsilon\, I(\Phi(\cdot - \gamma) = 1$, so that $\sup_{\gamma} I(F(\cdot, \gamma)) \leq 1$.

To verify (D)(i), suppose that $\int w(x)f'(x,\gamma)\,dx = 0$ for all γ , $a < \gamma < b$. Since $f'(x, \gamma) = -(1 - \varepsilon)x\phi(x) - \varepsilon(x - \gamma)\phi(x - \gamma)$, we have

$$(3.18) \qquad (1 - \varepsilon)\int xw(x)e^{-\frac{x^2}{2}}\,dx + \varepsilon\, e^{-\frac{\gamma^2}{2}} \int x\, w(x)e^{-\frac{x^2}{2}+\gamma x}dx - \varepsilon\gamma e^{-\frac{\gamma^2}{2}} \int w(x)e^{-\frac{x^2}{2}+\gamma x}dx = 0$$

for all $\gamma \in (a, b)$. Thus the region of convergence of the Laplace transform $\int xw(x)\, e^{-\frac{x^2}{2}}\, e^{sx}dx$ is an interval of s 's that includes both (a, b) and 0 . Setting $\gamma = 0$, we get $\int x\, w(x)\, e^{-\frac{x^2}{2}}\, dx = 0$ and substituting into (3.18), we obtain

$$(3.19) \qquad \int x\, w(x)\, e^{-\frac{x^2}{2}} e^{\gamma x}\, dx = \gamma \int w(x)\, e^{-\frac{x^2}{2}+\gamma x}\, dx$$

for all $\gamma \in (a, b)$. Since the left side of (3.19) is the Laplace transform of $x\, w(x)\, e^{-\frac{x^2}{2}}$ at γ and the right side is the Laplace transform of $-\frac{d}{dx}[w(x)e^{-\frac{x^2}{2}}]$ at γ, we conclude by the uniqueness theorem that

$$x\, w(x)\, e^{-\frac{x^2}{2}} = -\frac{d}{dx}[w(x)e^{-\frac{x^2}{2}}] = [x\, w(x) - w'(x)]\, e^{-\frac{x^2}{2}} \quad \text{a.e.} \tag{3.20}$$

so that $w'(x) = 0$ a.e. and $w(x) = C$ a.e. as desired. So the conclusions of theorem 1 hold.

To prove that F_0 is symmetric, note that $I(F(\cdot, \nu)) = I(F(\cdot, -\nu))$ for all ν, where $-\nu$ is the reflection of ν about 0. Then $F_0 = F(\cdot, \nu_0)$ could not be the unique distribution minimizing $I(\cdot)$ unless $\nu_0 = -\nu_0$.

As in the proof of theorem 2, the support of ν_0 is denumerable since, if not, $\psi_0 = -f_0'/f_0$ can be seen to be approximately linear in the tails so that $\sup_F V(\psi_0, F) = \infty$. Also since $\nu_0 = -\nu_0$, the accumulation points of ν_0 are at both $+\infty$ and $-\infty$.

It remains to show that $\nu_0\{-\infty\} = \nu_0\{+\infty\} = 0$. Arguing as in the proof of theorem 2, we need only show that

$$\int[-2\psi_0 g' - \psi_0^2 g]\, dx \geq 0 \qquad \text{for all} \quad g(x) = \int\phi(x-\gamma)\nu(d\gamma) \tag{3.21}$$

implies that $\psi_0 = 0$ a.e. Consider the following identity (which is straightforward to verify)

$$\int[2\psi_0(x)(x-\theta)\phi(x-\theta) - \psi_0^2\phi(x-\theta)]dx = 1 - E_\theta(X - \psi_0(X) - \theta)^2 , \tag{3.22}$$

where E_θ denotes expectation when the random variable X has a $N(\theta,1)$ distribution. Then (3.21) implies (by the choice ν = unit mass at θ) that the left side of (3.22) is $\geqslant 0$ for all θ, so that $\sup_\theta E_\theta[X - \psi_0(X) - \theta]^2 \leqslant 1$. But it is well known that X is the unique minimax estimator of θ, i.e., that if $\delta(X)$ is any other estimator of θ, then $\sup_\theta E_\theta[\delta(X) - \theta]^2 > 1$. So $\sup_\theta E_\theta[X - \psi_0(X) - \theta]^2 \leqslant 1$ implies that $\psi_0 \equiv 0$ a.e., completing the proof.

□

(iii) $$F(x, \gamma) = \Phi(x - \gamma)$$

Verification of conditions for the example is of course trivial if E is a proper interval. In that case the support of ν_0 is necessarily finite (as has long been known), and of course, ν_0 is symmetric about the midpoint of E . If E has even one infinite end point, $\inf \{I(F) : F \in \mathcal{F}\} = 0$ and Condition (A.7) fails. The assertion of theorem 1 on the nature of ν_0 continues to hold. However, ν_0 is not unique.

Acknowledgement

Dr. Marazzi (private communication) independently arrived at the heuristic argument behind theorem 3.

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press.

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *J. Royal Statist. Soc. Ser. B 36*, 99-102.

Bickel, P.J. (1980). Minimax estimation of the mean of a normal distribution subject to doing well at a point. Technical Report.

Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist. 9*, 1301-1309.

Casella, G. and Strawderman, W. (1981). Estimating a bounded normal mean. *Ann. Statist. 9*, 868-876.

Donoho, D.L. (1978). The asymptotic variance formula and large-sample criteria for the design of robust estimators. Unpublished Senior Thesis, Department of Statistics, Princeton University.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators - Part I: The Bayes case. *J. Amer. Statist. Assoc. 66*, 807-815.

Efron, B. and Olshen, R.A. (1978). How broad is the class of normal scale mixtures? *Ann. Statist. 6*, 1159-1164.

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

Huber, P.J. (1977). *Robust Statistical Procedures*. S.I.A.M. Philadelphia, PA.

Levit, B.Ya. (1979). On the theory of the asymptotic minimax property of second order. *Theory Probab. Appl. 24*, 435-437.

Levit, B.Ya. (1980). On asymptotic minimax estimates of the second order. *Theory Probab. Appl. 25*, 552-568.

Mallows, C.L. (1978). Problem 78.4, minimizing an integral. *S.I.A.M. Review 20*, 185

Mallows, C.L. (1980). Some theory of non-linear smoothers. *Ann. Statist. 8*, 695-715

Marazzi, A. (1980). Robust Bayesian estimation for the linear model. Research Report No. 27. Fachgruppe Fuer Statistik. E.T.H., Zurich.

Wu, C.F. (1978). Some algorithmic aspects of the theory of optimal design. *Ann. Statist. 6*, 1286-1301