Robust weighted Cramér–von Mises Estimators of location, with minimax variance in ε-contamination neighbourhoods

Douglas P. WIENS*

University of Alberta

Key words and phrases: Weighted Cramér–von Mises estimation, minimum-distance estimators, robust methods, minimax variance, ε-contamination, gross errors.


ABSTRACT

We construct weighted Cramér–von Mises location estimators which are asymptotically normally distributed throughout an ε-contamination neighbourhood of a given, strongly unimodal distribution function, and which minimize the maximum asymptotic variance in such neighbourhoods. Applications to the estimation of a normal or logistic mean are given.

RÉSUMÉ

On construit des estimateurs de position de Cramér-von Mises pondérés qui sont asymptotiquement distribués selon une loi normale partout un voisinage de contamination-ε d’une fonction de densité fortement unimodale donnée et qui minimisent la variance asymptotique maximum dans de tels voisinages. On applique ces résultats à l’estimation de la moyenne d’une fonction de densité normale ou logistique.

1. INTRODUCTION

We consider weighted Cramér–von Mises estimation of a location parameter, with an eye to constructing robust, asymptotically normal estimators which enjoy minimax variance properties in ε-contamination (“gross errors”) neighbourhoods of the “target” distribution. The estimators are members of the general class of minimum-distance estimators first studied by Wolfowitz (1957). Parr (1981) gives an extensive bibliography of papers related to minimum-distance estimation. In particular, Millar (1981) discusses weighted Cramér–von Mises estimation from a decision-theoretic viewpoint, and determines the types of contamination against which a specified weight function offers protection. Boos (1981) and Parr and DeWet (1981) determine weight functions which are asymptotically efficient when the target distribution is correctly specified.

Here, we specify a contamination neighbourhood by assuming that the sample values arise through possible ε-contamination of a known distribution G. Under

---

*Research supported by the Natural Sciences and Engineering Research Council of Canada, under grant A-8603.
some regularity conditions on $G$, the most important of which is strong unimodality, we show that a suitably truncated version of the weight function efficient for $G$ yields an estimator which is asymptotically normal for all distributions in such a neighbourhood, and whose maximum asymptotic variance, as the true distribution varies over this neighbourhood, is a minimum. The results, with examples, are discussed in Section 2 and proven in Section 3.

2. STATEMENT OF MAIN RESULTS

Given a random sample $X_1, \ldots, X_n$ with distribution function $F(x - \theta)$, and a specified distribution function $F_0$, not necessarily equal to $F$, we consider estimators $\theta_n$ which result from minimizing

$$d_{F_n}(\theta) = \int_{-a}^{b} (F_n(x) - F_0(x - \theta))^2 w(x - \theta) \, dx,$$

where $F_n$ is the d.f. of the sample. It is assumed that:

(A1) In $[-a, b]$, $w(x)$ is nonnegative, absolutely continuous with respect to Lebesgue measure, and with a continuous derivative $w'(x)$ (one-sided, at $-a$ and $b$); and $w(x) = 0$ off of $[-a, b]$.

(A2) $F, F_0$ have continuous densities $f, f_0$ w.r.t. Lebesgue measure.

Then $d_{F_n}(\theta)$ and $d_F(\theta)$ have derivatives $d'_{F_n} = \lambda_{F_n}, d'_{F} = \lambda_{F}$ given by

$$\lambda_{F_n}(\theta) = \frac{2}{n} \sum_{i=1}^{n} \left[ \frac{i-0.5}{n} - F_0(X_{(i)} - \theta) \right] w(X_{(i)} - \theta),$$

(2.2)

where $X_{(i)}$ is the $i$th-order statistic, and

$$\lambda_{F}(\theta) = w(b) (F(b - \theta) - F_0(b))^2 - w(0) (F(a - \theta) - F_0(a))^2$$

$$+ 2 \int_{-a}^{b} (F(x + \theta) - F_0(x)) f_0(x) w(x) \, dx$$

$$- \int_{-a}^{b} (F(x + \theta) - F_0(x))^2 w'(x) \, dx.$$  

(2.3)

As in Parr and DeWet (1981), we take the estimand $\theta(F)$ to be a zero of $\lambda_{F}(\theta)$, assume the existence of a consistent estimator $\hat{\theta}$ of $\theta$, and define the estimator $\theta_n = \hat{\theta}(F_n)$ to be the closest zero, to $\hat{\theta}$, of $\lambda_{F_n}(\theta)$. In practice, one would use $(X - \theta)/\sigma$ in (2.1) and minimize over both $\theta$ and $\sigma$ to obtain an auxiliary estimate of scale. Since the two estimators so obtained are typically asymptotically independent (Parr and Schucany 1980) under the symmetry assumptions made below, and we are here primarily interested in establishing the minimax property of the location estimate, we shall assume that scale is known.

The influence curve (Hampel 1974)

$$IC(z; F) = \lim_{s \to 0} \frac{\theta((1 - s)F + s \delta_z) - \theta(F)}{s},$$

where $\delta_z$ is unit mass at $z$, of $\theta_n$ is obtained by implicit differentiation of the relationship $\lambda_{F}(\theta) = 0$, and is found to be
\[
IC(z; F) = \left( \frac{1}{2} \lambda_F'(\theta(F)) \right)^{-1} \\
\times (w(-a)(F(\theta - a) - F_0(-a))(I(z \leq \theta - a) - F(\theta - a)) \\
- w(b)(F(\theta + b) - F_0(b))(I(z \leq \theta + b) - F(\theta + b)) \\
+ \int_{-a}^{b} [I(z \leq x + \theta) - F(x + \theta)] \\
\times [(F(x + \theta) - F_0(x))w'(x) - f_0(x)w(x)] \, dx).
\] (2.4)

**Theorem 1** (Asymptotic normality at \( F \)). If \( \lambda_F'(\theta(F)) \neq 0 \) then

\[
\sqrt{n}(\hat{\theta}_n - \theta) \overset{D}{\rightarrow} N(0, E_F[IC^2(X; F)]).
\]

Note that IC(z; F) is bounded in \( z \), and constant in the tails, but that it may be discontinuous at \( \theta + b \) [\( \theta - a \)] if \( F_0(b) [F_0(-a)] \) is incorrectly specified.

Now assume that both \( F \) and \( F_0 \) are symmetric about \( \theta \), which we take without loss of generality to be zero, and put \( b = a \). The initial estimator \( \hat{\theta} \) may be taken to be the sample median, trimmed mean, or any other reasonably robust estimator of a centre of symmetry. Now (2.4) becomes IC(z; F) = \( A_F(z)/B_F \), where

\[
A_F(z) = \int_{-a}^{z} f(x)w(x) \, dx - (F - F_0)(z)w(z),
\] (2.5)

\[
B_F = 2w(a)f(a)(F - F_0)(a) + \int_{-a}^{a} f(x)f_0(x)w(x) \, dx \\
- \int_{-a}^{a} (F - F_0)(x)f(x)w'(x) \, dx.
\] (2.6)

Assume further that \( F \) is an arbitrary member of a gross-errors neighbourhood \( \mathcal{G}_{\varepsilon} \) of a known distribution function \( G \):

\[
\mathcal{G}_{\varepsilon} = (F | F' = f = (1 - \varepsilon)g + \varepsilon h; h \text{ symmetric and continuous}).
\]

Of \( G \) we assume:

(G1) \( G' = g \) is symmetric and continuous.

(G2) The score function \( \xi = -g'/g \) is strictly increasing and twice continuously differentiable.

Choose \( F_0 \) to have minimum Fisher information for location, \( I(F) = \int (f'/f)^2f \, dx \), in \( \mathcal{G}_{\varepsilon} \). Huber (1964) showed that \( F_0 \) is given by

\[
f_0(x) = \begin{cases} 
(1 - \varepsilon)g(x), & |x| \leq a, \\
(1 - \varepsilon)g(a)\exp(-\xi(a)(|x| - a)), & |x| \geq a,
\end{cases}
\]

with \( a \) and \( \varepsilon \) related by \( \int f_0 = 1 \), i.e.

\[
\frac{\varepsilon}{2(1 - \varepsilon)} = \frac{g(a)}{\xi(a)} - \bar{G}(a).
\] (2.7)

Here and throughout, we use the notation \( \bar{G} = 1 - G \).

Any \( F \in G_{\varepsilon} \) satisfies (A2). Denote by \( W \) the set of weight functions satisfying (A1) [with \( b = a \) determined from (2.7)], and for which \( \inf_{G_{\varepsilon}} B_F > 0 \). Note that since
\( f \geq f_0 > 0 \) in \([-a, a]\), \( I(|x| \leq a) \in W \) and so \( W \neq \emptyset \). Denote by \( V(w, F) \) the asymptotic variance of \( \sqrt{n}(\theta_n - \theta) \), if \((w, F) \in W \times \mathcal{G}_c\). Then \( w_0 \in W \) satisfies the minimax property if

\[
\inf_{W} \sup_{\mathcal{G}_c} V(w, F) = \sup_{\mathcal{G}_c} V(w_0, F). \tag{2.8}
\]

See Huber (1981), Jaeckel (1971), Collins (1983) for classes of estimators and neighbourhoods for which the minimax property is known to hold; Sacks and Ylvisaker (1972, 1982), Collins and Wiens (1986) for classes in which it does not. We adopt the same approach as those authors above, who obtained positive results. That is, we isolate a weight function \( w_0 \) which yields an estimator which is efficient at \( F_0 \), and then seek to verify the saddlepoint property

\[
V(w_0, F) \leq V(w_0, F_0) = \frac{1}{I(F_0)} \leq V(w, F_0) \quad \text{for all } (w, F) \in W \times \mathcal{G}_c, \tag{2.9}
\]

which implies (2.8). From Boos (1981) and Parr and DeWet (1981), it is seen that, with \( \psi_0 = -f_0'/f_0 \), the proper choice is (any nonzero multiple of)

\[
w_0(x) = \frac{\psi_0(x)}{f_0(x)} = \frac{\xi'(x)}{(1 - \varepsilon)g(x)} I(|x| \leq a),
\]

which satisfies (A1). If \( w_0 \in W \), then

\[
V(w_0, F) := V_0(F) = 2 \int_0^\infty \frac{A_F(z)}{B_F} \frac{dF(z)}{B_F},
\]

where, after an integration by parts in (2.5),

\[
A_F(z) = \begin{cases} 
\psi_0(z) - \int_0^z (F - F_0)(x)w_0'(x) \, dx, & 0 \leq z \leq a, \\
A_F(a) + w_0(a)(F - F_0)(a), & z > a; 
\end{cases} \tag{2.10}
\]

\[
B_F = 2w_0(a)f(a)(F - F_0)(a) + 2 \int_0^a f(x)\psi_0(x) \, dx \\
- 2 \int_0^a (F - F_0)(x)f(x)w_0'(x) \, dx. \tag{2.11}
\]

Note that \( A_F(z) = \psi_0(z), \ B_{F_0} = I(F_0) > 0, \ V_0(F_0) = 1/I(F_0) \). The second inequality in (2.9) is essentially the Cauchy-Schwartz inequality. With \( \psi(z) = \int_0^z f_0(x)w(x) \, dx \) in (2.5), (2.6), we have

\[
V(w, F_0) = \frac{\int \psi f_0}{\left( \int \psi f_0 \right)^2} \geq \frac{1}{\int (f_0/f_0)^2 f_0} = \frac{1}{I(F_0)} = V(w_0, F_0).
\]

Conditions under which \( w_0 \in W \) are given by
Theorem 2 (Asymptotic normality throughout $\mathcal{G}_e$). If $G$ is such that either

\begin{align*}
\text{(G3 i) } w_0'(x) &\leq 0, x \in [0, a] \text{ or} \\
\text{(G3 ii) } w_0'(x) &\geq 0, \xi''(x) \leq 0, x \in [0, a]
\end{align*}

then $w_0 \in W$, so that

$$\sqrt{n}(\theta_n - \theta)^{-1}N(0, V_0(F)) \quad \text{for all } F \in \mathcal{G}_e.$$

Example 1. Condition (G3 i) applies to the hyperbolic-secant distribution (the distribution of $\log |Z|$ if $Z$ is a Cauchy r.v.), for which $g(x) = \pi \sech x$ and $w_0(x) = g(x)I(|x| \leq a)/(1 - \varepsilon)$.

Theorem 3 (Attainment of the minimax property). Suppose that conditions (G1), (G2), (G3 ii) hold, and that in addition

\begin{itemize}
  \item[(G4')] $w_0$ is convex ($w_0'$ nondecreasing) in $[-a, a]$, or merely
  \item[(G4')] $\sup_{[0,a]} w_0(x) = w_0(a)$.
\end{itemize}

Then $\sup_{\mathcal{F}_e} V_0(F) = 1/I(F_0)$, so that the minimax property (2.8) holds.

Example 2. If $G = \Phi$ is the standard normal distribution function, then $w_0(x) = ((1 - \varepsilon)\Phi(x))^{-1}I(|x| \leq a)$ is minimax, with $\sup_{\mathcal{F}_e} V_0(F) = [(1 - \varepsilon)(1 - 2\Phi(a))]^{-1}$. See Huber (1981) for some numerical values of $a$ and $\varepsilon$. Parr and DeWet (1981) suggested the use of this weight function in estimating the mean of a normal distribution, apparently without being aware of its optimality. They recommended the trimmed mean, with trimming proportion $\Phi(a)$, as a starting value.

Example 3. For $G(x) = (1 + e^{-x})^{-1}$, the logistic distribution function, the weights $w_0(x) = 2I(|x| \leq a)/(1 - \varepsilon)$ are optimal. With $t = G(a) - \frac{1}{2}$, the relevant values are $\varepsilon = (1 - 2t)^2/(1 + 4t^2)$ and $\sup_{\mathcal{F}_e} V_0(F) = 3(1 + 4t^2)/(4t^2(3 - 4t^2))$. Since

$$w_0(x) \propto \frac{g(x)}{G(x)G'(x)}I(|x| \leq a),$$

$\theta_n$ can be heuristically described as an Anderson-Darling estimator, with truncated weights.

The generalized logistic density [Gumbel (1944); see also Johnson and Kotz (1970)] is defined by $g_m(x) = \text{const. } g^{m+1}(x)$, with $g(x)$ as above. In this case, the weights should be chosen proportional to $g^{-m}(x)I(|x| \leq a)$.

3. PROOFS

3.1. Proof of Theorem 1.

The proof parallels that of Theorem 3.1 of Boos (1981), who assumes a continuous weight function. Define

$$\Delta_n(x) = F_n(x) - F(x),$$

$$h_r(t) = \begin{cases} 
\frac{\lambda_r(t) - \lambda_r(0)}{2(t - \theta)}, & t \neq \theta(F), \\
\lambda_r(\theta(F)), & t = \theta(F);
\end{cases}$$

$$\Delta_{n_0}(x) = F_{n_0}(x) - F(x),$$

$$h_r(t) = \begin{cases} 
\frac{\lambda_r(t) - \lambda_r(0)}{2(t - \theta)}, & t \neq \theta(F), \\
\lambda_r(\theta(F)), & t = \theta(F);
\end{cases}$$

$$\Delta_{n_0}(x) = F_{n_0}(x) - F(x),$$

$$h_r(t) = \begin{cases} 
\frac{\lambda_r(t) - \lambda_r(0)}{2(t - \theta)}, & t \neq \theta(F), \\
\lambda_r(\theta(F)), & t = \theta(F);
\end{cases}$$
\[
T(F; \Delta_n) = (h_F(\theta))^{-1} \left( w(-a) \Delta_n(\theta - a)(F(\theta - a) - F_0(-a))
- w(b) \Delta_n(\theta + b)(F(\theta + b) - F_0(b))
+ \int_{\theta-a}^{\theta+b} \Delta_n(x) \left( (F(x) - F_0(x - \theta))w'(x - \theta) - f_0(x - \theta)w(x - \theta) \right) dx \right).
\]

Since \( \sqrt{n} T(F; \Delta_n) = n^{-1/2} \sum_{i=1}^{n} 1IC(X_i; F) \), it suffices to establish
\[
\theta_n - \theta = T(F; \Delta_n) + o_p(n^{-1/2}). \tag{3.1}
\]

The consistency of \( \hat{\theta} \) implies that of \( \theta_n \), so that
\[
h_F(\theta_n) \overset{P}{\rightarrow} h_F(\theta) \neq 0.
\]

Then (3.1) is a consequence of
\[
|h_F(\theta_n)| \left| \theta_n - \theta - \frac{h_F(\theta)}{h_F(\theta_n)} T(F; \Delta_n) \right|
= \frac{\lambda_F(\theta_n) - \lambda_{F,n}(\theta_n)}{2} - h_F(\theta)T(F; \Delta_n) = o_p(n^{-1/2}). \tag{3.2}
\]

Define \( ||\Delta_n||_a = \sup_{-\infty, x < \infty} \left| \frac{\Delta_n(x)}{[F(x)(1 - F(x))]^{1/2}} \right| \). Then the second term in (3.2) is bounded above by
\[
\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \Delta_n^2(x) |w'(x - \theta_n)| dx
+ \int_{-\infty}^{\infty} \Delta_n(x) |f_0(x - \theta)w(x - \theta) - f_0(x - \theta_n)w(x - \theta_n)| dx
+ \int_{-\infty}^{\infty} \Delta_n(x) |F_0(x - \theta) - F_0(x - \theta_n)| |w'(x - \theta_n)| dx
+ \int_{-\infty}^{\infty} \Delta_n(x) |F(x) - F_0(x - \theta)| |w'(x - \theta_n) - w'(x - \theta)| dx
+ \sum_{c \in (-\alpha, b)} w(c) \left[ \frac{1}{2} \Delta_n^2(\theta_n + c) + |\Delta_n(\theta_n + c)| |F_n(\theta_n + c) - F(\theta + c)|
+ |\Delta_n(\theta_n + c) - \Delta_n(\theta + c)| |F(\theta + c) - F(0)| \right] \right\}
\leq \frac{1}{2} ||\Delta_n||_{1 - \varepsilon} ||\Delta_n||_{1 - \varepsilon} \int_{-\infty}^{\infty} [F(x)(1 - F(x))]^{1/2} |w'(x - \theta_n)| dx
+ ||\Delta_n||_{1 - \varepsilon} \int_{-\infty}^{\infty} [F(x)(1 - F(x))]^{1/2} |f_0(x - \theta)w(x - \theta) - f_0(x - \theta_n)w(x - \theta_n)| dx
+ ||\Delta_n||_{1 - \varepsilon} \int_{-\infty}^{\infty} F(x)(1 - F(x))]^{1/2} |F_0(x - \theta) - F_0(x - \theta_n)| |w'(x - \theta_n)| dx
+ |F(x) - F_0(x - \theta)| |w'(x - \theta_n) - w'(x - \theta)| dx
\[ + \sum_{c \in (-a, b)} w(c) \left[ \frac{1}{2} \| \Delta_n \|_{2} - \epsilon \| \Delta_n \|_1 - \epsilon [F(\theta + c)(1 - F(\theta + c)) - F(\theta + c)] \|_{2} - \epsilon |F(\theta + c) - F(\theta + c)| \right] \] 
\[ + | \Delta_n(\theta + c) - \Delta_n(\theta + c) | \right] \]

As pointed out by Boos (1981), it follows from Lemmas 7.1, 7.3 of Gregory (1977) that
\[ ||\Delta_n||_{1-\epsilon} = O_p(n^{-1/2}) \text{ and } ||\Delta_n||_{1-\epsilon} - 0, \quad \text{for} \quad \epsilon > 0. \]

We can represent \(| \Delta_n(\theta + c) - \Delta_n(\theta + c) |\) as \(n^{-1} |Y_n - np_n|\), where \(Y_n \sim \text{bin}(n, p_n)\) and \(p_n = |F(\theta + c) - F(\theta + c)| \rightarrow 0\). It follows from these observations that the term at (3.3) is \(o_p(n^{-1/2})\).

### 3.2. Proof of Theorem 2.

We need only verify that \(\inf_{\psi} B_F > 0\) under either condition. From (2.11),
\[ B_F \geq 2 \int_0^a [\psi_0' - (F - F_0)w_0'] f \, dx. \]
If (G3 i) holds, then
\[ B_F \geq 2 \int_0^a \psi_0' f_0 \, dx = I(F_0) > 0. \]
If (G3 ii) holds, then (2.7) gives
\[ (\psi_0 - \epsilon \psi)_0' = \frac{\xi(x)}{g(x)} \left( g(x) \frac{g'(a)}{g(x)} - g(a) \frac{g'(a)}{g(x)} \right) + (1 - \epsilon)w_0(x)\overline{G}(a) > 0, \]
so that
\[ B_F \geq 2 \int_0^a (\psi_0 - \epsilon \psi)_0' f \, dx \geq 2 \int_0^a (\psi_0 - \epsilon \psi)_0' f_0 \, dx > 0. \]

### 3.3 Proof of Theorem 3.

We repeatedly use the relationship
\[ \psi_0(a) = \frac{f_0(a)}{F_0(a)}, \]
which is equivalent to (2.7). The first step in the proof is to bound \(V_{\psi}(F)\) by a more tractable functional \(V_{\psi_k}(F)\), for each \(k = 2\epsilon^{-1}(F - F_0)(a) \in [0, 1]\). Define
\[ \mathcal{F}_k = (F \mid F = (1 - \epsilon)G + \epsilon H, H \text{ a symmetric distribution function} \]
with \(H([-a, a]) = k\), \(k \in [0, 1]\).

Note that \(\mathcal{F}_k\) is convex, and that \(\mathcal{G}_k \subseteq \bigcup_k \mathcal{F}_k\), since we do not require \(H\) to possess a density. For \(F \in \mathcal{F}_k\), define
\[ V_{\psi_k}(F) = \frac{2 \left[ \int_0^a \psi_0 \, dF + \left( \psi_0(a) + \frac{\epsilon k}{2} w_0(a) \right)^2 \left( F_0(a) - \frac{\epsilon k}{2} \right) \right]}{\left[ \int_0^a (4\psi_0 + \psi_0^2) \, dF - \left( w_0(a) \frac{\epsilon k}{2} + \psi_0(a) \frac{\epsilon k}{2} + \int_0^a (2\psi_0 + \psi_0^2) \, dF_0 \right) \right]^2}. \]
LEMMA. If \( F \in \mathcal{G}_e \cap F_k \), then \( V_0(F) \leq V_{(k)}(F) \).

Proof. Using (G3 ii) in (2.10),

\[
0 \leq A_F(z) \leq A_F(z) := \begin{cases} 
\psi_0(z), & 0 \leq z \leq a, \\
\psi_0(a) + \frac{\varepsilon k}{2} w_0(a), & a < z.
\end{cases}
\]

In (2.11), write \( \int_0^a (F - F_0)w_0^f \, dx \) as

\[
\frac{1}{2} \int_0^a w_0'(x) \frac{d}{dx} (F - F_0)^2(x) \, dx + \int_0^a (F - F_0)(x) \frac{d}{dx} (\psi_0'(x) + \frac{1}{2} \psi_0'(x)) \, dx
\]

and integrate by parts, obtaining

\[
B_F = \left\{ \int_0^a (4\psi_0' + \psi_0) \, d(F - F_0) \\
+ 2 \int_0^a \psi_0' \, dF_0 - w_0'(a)(F - F_0)^2(a) - \psi_0'(a)(F - F_0)(a) \right\}
\]

\[
+ \left( \int_0^a (F - F_0)^2 \, dw_0' + 2w_0(a)(F - f_0)(a)(F - F_0)(a) \right).
\]

Denote by \( B_F^* \) the term in braces. Either of (G4) or (G4') implies that \( B_F^* \geq B_F^* \), and then (G3 ii) gives

\[
B_F^* \geq 4\psi_0'(a)(F - F_0)(a) + 2f_0(a)\psi_0(a) - w_0'(a)(F - F_0)^2(a) - \psi_0'(a)(F - F_0)(a).
\]

Since \( F \in \mathcal{F}_k \), this last term may be written as \((\varepsilon/2)p(k)\), where the quadratic

\[
p(k) = -\frac{\varepsilon}{2} k^2 w_0'(a) + k(4\xi' - \xi^2)(a) + \frac{4(1 - \varepsilon)}{\varepsilon} g(a)\xi(a)
\]

is clearly positive at \( k = 0 \), and can be seen, using (2.7) and \( w_0'(a) \leq \psi_0(a)w_0(a) \), to be positive at \( k = 1 \). It is thus positive throughout \([0, 1]\), and so \( B_F \geq B_F^* > 0 \). Then

\[
V_0(F) \leq 2 \int_0^\infty \frac{A_F^2 \, dF}{B_F^2} = V_{(k)}(F).
\]

Q.E.D.

It is now sufficient to show that \( \sup_{\mathcal{G}_e} V_{(k)}(F) \leq 1/I(F_0) \) for each \( k \in [0, 1] \). If \( k = 0 \), then \( V_{(k)}(F) = 1/I(F_0) \). If \( k > 0 \), then \( \mathcal{F}_k \) may be identified in an obvious way with the set of all distribution functions on \([0, 1]\), and is thus weakly compact by Prohorov's theorem. Since \( V_{(k)}(F) \) is weakly continuous, it attains its supremum over \( \mathcal{F}_k \), at \( F_k \), say. By Lemma 4.4 of Huber (1981), \( \varphi(t) = (V_{(k)}((1 - t)F_k + tF_1))^{-1} \) is a convex function of \( t \in [0, 1] \) for each \( F_i \in \mathcal{F}_k \); hence \( (V_{(k)}(F))^{-1} \) is minimized by \( F_k \) iff \( \varphi'(0) \geq 0 \) for each \( F_i \), i.e. iff

\[
0 \leq \int_0^a (4\psi_0'(x) + (1 - r_k)\psi_0'(x)) \, d(F_1 - F_k), \quad \text{all } F_i \in \mathcal{F}_k,
\]

where \( r_k = (V_{(k)}(F_k))^{-1} \). Within \([0, a]\), \( F_k \) is then necessarily of the form

\[
\psi_0(x) + \frac{\varepsilon k}{2} w_0(x),
\]

for some \( \psi_0(x) \) and \( \varepsilon k \).
\[ F_k = F_0 + \frac{\varepsilon k}{2} \delta_c, \]  
\tag{3.7}

where \( \delta_c \) is unit mass at \( c = c(k) \), and \( c \) is any point at which \( 4\psi' + (1 - r_k)\psi_0 \) has a minimum in \([0, a]\). Otherwise, choosing \( F_1 \) to be of the form (3.7) would violate (3.6). Rather than determine \( c(k) \) explicitly—this can be done in specific cases—we shall show that

\[ V_{(k)}(F_k) \leq \frac{1}{I(F_0)} \]  
\tag{3.8}

for all distributions of the form (3.7). For such distributions,

\[ V_{(k)}(F_k) = \frac{I(F_0) + T_1}{(I(F_0) + T_2)^2}, \]

where

\[ T_1 = \varepsilon k [\psi_0(c) - \psi_0(a) + 2\psi_0(a)] + (\varepsilon k)^2 \frac{w_0}{2\psi_0}(a)[\psi_0(a) - 2\psi_0(a)] - \frac{(\varepsilon k)^3}{4} \psi_0(a), \]

\[ T_2 = \frac{\varepsilon k}{2} \left[ 4\psi_0(c) + \psi_0(c) - \psi_0(a) \right] - \frac{(\varepsilon k)^2}{4} w_0(a). \]

Then (3.8) becomes

\[ T_2^2 + 2T_2 - T_1I(F_0) \geq 0. \]  
\tag{3.9}

Define

\[ R = \frac{\psi_0(a)}{\psi_0(a)} > 0, \quad \beta = 2\psi_0(c) - \psi_0(a) \geq R\psi_0(a), \]

\[ \alpha = R(R - 1) \frac{\psi_0^3}{f_0}(a) + \frac{\psi_0''}{f_0}(a) \leq R(R - 1) \frac{\psi_0^3}{f_0}(a). \]

Then \( 2T_2 - T_1 = \varepsilon k[(\varepsilon k)^2 w_0(a) - 2\varepsilon k + 8\beta]/4. \) We can assume that this is negative, so that the quadratic, as a function of \( \varepsilon k \), must have two real roots. This implies \( R > 1 \) and \( R^2 - 10R + 1 > 0 \), i.e.

\[ R > 5 + 2\sqrt{6}. \]  
\tag{3.10}

It follows from (3.10) that the function \( \frac{1}{4}(\sqrt{6} - 2)\xi g(x) - \int_0^x \xi^2 g \, dt \) is increasing in \([0, a]\), so that

\[ S := \frac{I(F_0)}{\psi_0(a)f_0(a)} = 2 + \frac{2}{\xi(a)g(a)} \int_0^a \xi^2 g \, dt \in \left(2, \frac{2 + \sqrt{6}}{2}\right). \]  
\tag{3.11}

With

\[ \gamma = 4\psi_0(c) + \psi_0(c) - \psi_0(a) > 4\psi_0(a) - \psi_0(a) = (4R - 1)\psi_0(a), \]  
\tag{3.12}

\[ A = \gamma^2 - 2R(R - 1)S\psi_0(a) - 2S\psi_0(a) \psi_0(a), \]
\[ B = R\psi_0'(a) \frac{RS\psi_0(a) - \gamma}{\tilde{F}_0(a)} - \frac{\gamma\psi_0''(a)}{f_0(a)}, \]

the inequality (3.9) becomes

\[ 0 \leq \left( \frac{w_0'(a)}{4} \right)^2 (\varepsilon k)^3 + 2\beta I(F_0) + \frac{\varepsilon k(A + B\varepsilon k)}{4}. \]

This is satisfied by all \( \varepsilon k \in [0, 1] \) if both \( A \) and \( A + B \) are positive. That \( A > 0 \) follows from (3.11), (3.12). Then using as well \( \tilde{F}_0(a) \leq \frac{1}{2} \),

\[ A + B \geq (\gamma - R\psi_0'(a))^2 + (2S - R)R\psi_0'(a) > (8R^2 - 2R + 1)\psi_0'(a) > 0. \]

REFERENCES


