ASYMPTOTIC MINIMAX PROPERTIES OF M-ESTIMATORS OF SCALE

Douglas P. WIENS * and K.H. Eden WU

Department of Statistics and Applied Probability, University of Alberta, Edmonton, Canada T6G 2G1

Abstract: We ask whether or not the saddlepoint property holds, for robust M-estimation of scale, in gross-errors and Kolmogorov neighbourhoods of certain distributions. This is of interest since the saddlepoint property implies the minimax property — that the supremum of the asymptotic variance of an M-estimator is minimized by the maximum likelihood estimator for that member of the distributional class with minimum Fisher information. Our findings are exclusively negative — the saddlepoint property fails in all cases investigated.

1. Introduction and summary

We consider the problem of minimax, robust M-estimation of scale, when the distribution generating the observations is only partially known. To set the framework for the problem, recall that an M-estimate of scale is defined as $S(F_n)$, where F_n is the empirical distribution function based on a sample $X_1, \ldots, X_n \sim F$, and the functional $S(F)$ is defined implicitly by

$$\int_{-\infty}^{\infty} \chi\left(\frac{x}{S(F)} \right) \, dF(x) = 0. \tag{1.1}$$

The function χ is chosen by the statistician. Under appropriate regularity conditions (see, for example, Boos and Serfling, 1980; Serfling, 1980),

$$\sqrt{n}\left(\frac{S(F_n)}{S(F)} - 1 \right) \xrightarrow{w} N(0, V(\chi, F)), \tag{1.2}$$

where the *standardized asymptotic variance* (Bickel and Lehmann, 1976) is given by

$$V(\chi, F) = \int_{-\infty}^{\infty} \chi^2\left(\frac{x}{S(F)} \right) dF(x) \Big/ \left[\int_{-\infty}^{\infty} \chi'\left(\frac{x}{S(F)} \right) \frac{x}{S(F)} \, dF(x) \right]^2. \tag{1.3}$$

Suppose that F is known only to belong to a certain convex class $\mathscr{F}$ of distributions. Two such classes, common in the robustness literature, are the *gross-errors neighbourhood*

$$\mathscr{G}_\varepsilon(G) = \{ F = (1 - \varepsilon)G + \varepsilon H, \ H \text{ arbitrary} \},$$

and the *Kolmogorov neighbourhood*

$$\mathcal{K}_\varepsilon(G) = \left\{ F \mid \sup_{-\infty < x < \infty} |F(x) - G(x)| \le \varepsilon \right\},$$

with ε and G known.

For $F(= F_{(1)}) \in \mathcal{F}$, define $F_{(\sigma)}$ by

$$F_{(\sigma)}(x) = F(x/\sigma), \quad \sigma > 0.$$

Define Fisher information for scale by

(a) $\qquad I(F; \sigma) = I(F_{(\sigma)}; 1) = \dfrac{1}{\sigma^2} I(F; 1);$

(b) $\qquad I(F; 1) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \left(-x \dfrac{f'}{f}(x) - 1 \right)^2 f(x)\,\mathrm{d}x & \text{if } F \text{ has a density } f, \text{ absolutely continuous on} \\ & \mathbb{R} \setminus \{0\}, \\ \infty & \text{otherwise.} \end{cases}$

Suppose that F_0 minimizes $I(F; 1)$ in $\mathcal{F}$. Define $\chi_0(x) = -x(f_0'/f_0)(x) - 1$, corresponding to maximum likelihood estimation of σ if $X_1, \ldots, X_n \sim F_{0,(\sigma)}$. Define $S_0(F)$ by (1.1), with $\chi = \chi_0$. Then $S_0(F_0) = 1$. We have

$$V(\chi_0, F) \le V(\chi_0, F_0) = 1/I(F_0; 1) \le V(\chi, F_0) \tag{1.4}$$

for all $F \in \mathcal{F}_1 = \{ F \in \mathcal{F} \mid S_0(F) = 1 \}$ and all χ such that (1.1) holds for $F \in \mathcal{F}_1$. The second inequality in (1.4) is essentially the Cramér–Rao Inequality; the first is established by variational arguments, as in Huber (1964, 1981). It follows from (1.4) that

$$\sup_{\mathcal{F}_1} V(\chi_0, F) = \inf_{\chi} \sup_{\mathcal{F}_1} V(\chi, F), \tag{1.5}$$

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

In this paper, we address the question of whether or not (1.4) extends from $\mathcal{F}_1$ to all of $\mathcal{F}$, for $\mathcal{F} = \mathcal{G}_\varepsilon(G)$ and $\mathcal{F} = \mathcal{K}_\varepsilon(G)$ and with G satisfying certain assumptions. If so, i.e. if

$$\sup_{\mathcal{F}} V(\chi_0, F) = V(\chi_0, F_0), \tag{1.6}$$

we say that the *saddlepoint property* holds.

There is a fairly broad literature concerning the problem of whether or not various robust estimators possess the saddlepoint property. This literature concentrates on *location* problems almost exclusively — see Huber (1964, 1981), Collins (1983), Collins and Wiens (1985, 1989), Sacks and Ylvisaker (1972, 1982), Wiens (1986, 1987). For *scale* estimation the only reported results on this problem are those of Huber (1981), who verifies (numerically) that the saddlepoint property holds for the class $\mathcal{G}_\varepsilon(\Phi)$, with Φ the standard normal d.f. and $\varepsilon \le 0.04$.

In contrast to the positive result of Huber (1981), we show in Section 2 of this paper that the saddlepoint property (1.6) *fails* for scale estimation in:

(i) $\mathcal{G}_\varepsilon(G)$, when the 'large ε' form of the solution (see (2.1) below) applies;

(ii) $\mathcal{G}_\varepsilon(\Phi)$, $0.0997 \le \varepsilon \le 0.2051$ (= the boundary between the 'small ε' and 'large ε' forms, if $G = \Phi$);

(iii) $\mathcal{K}_\varepsilon(G)$, all ε.

In each case we make some mild assumptions on G. In all cases we assume that G is symmetric. We do not assume that the other members of $\mathcal{F}$ are symmetric.

Note that our negative results do not mean that (1.5) fails, if $\mathscr{F}_1$ is replaced by $\mathscr{F}$. To verify (1.5), however, without the benefit of (1.6), would require one to perform the, evidently very intractable, task of determining $\sup_{\mathscr{F}} V(\chi, F)$ for each fixed χ, and to then determine the infimum thereof.

2. Saddlepoint results

2.1. $\mathscr{G}_\varepsilon(G)$, large ε

Assume that G is symmetric, with $I(G; 1) < \infty$, and that the density $g_*(x) = 2e^x g(e^x)$ of $\log |X|$, when $X \sim G$, is strongly unimodal. Wu (1990) noted that the latter is a very mild restriction on G. Typical distributions G satisfying these conditions are the normal, logistic, Student's t and Laplace.

Under these conditions, Huber (1981) showed that there is an $\varepsilon_0(G)$ such that for $\varepsilon > \varepsilon_0$, $I(F; 1)$ is minimized over $\mathscr{G}_\varepsilon(G)$ by that F_0 for which

$$-x\frac{f_0'}{f_0}(x) - 1 = \chi_0(x) = \begin{cases} -k, & |x| \leqslant x_0, \\ \xi(x), & x_0 \leqslant |x| \leqslant x_1, \\ k, & |x| \geqslant x_1. \end{cases} \qquad (2.1)$$

Here,

$$\xi(x) = -x\frac{g'}{g}(x) - 1.$$

The constants k, x_0, x_1 are determined by the requirement that χ_0 be continuous, and that f_0 have mass one. It turns out that x_0 is an increasing function of ε, and then $\varepsilon_0(G)$ is defined by '$x_0 = 0$'. If $G = \Phi$, then $\varepsilon_0 = 0.2051$.

For $F_1 \in \mathscr{G}_\varepsilon(G)$, put

$$F_t = (1 - t)F_0 + tF_1, \quad 0 \leqslant t \leqslant 1, \qquad u(t) = 1/V(\chi_0, F_t).$$

Consider the following subsets of $\mathscr{G}_\varepsilon(G)$:

$$\mathscr{G}_\varepsilon^1(G) = \left\{ F = (1 - \varepsilon)G + \varepsilon H \mid I(F; 1) < \infty, H\{[-x_0, x_0]\} = 1 \right\},$$

$$\mathscr{G}_\varepsilon^2(G) = \left\{ F = (1 - \varepsilon)G + \varepsilon H \mid I(F; 1) < \infty, H\{(-\infty, -x_1] \cup [x_1, \infty)\} = 1 \right\},$$

$$\mathscr{G}_\varepsilon^*(G) = \mathscr{G}_\varepsilon^1(G) \cup \mathscr{G}_\varepsilon^2(G).$$

We will show that if $F_1 \in \mathscr{G}_\varepsilon^*(G)$, then

$$u'(0) = 0, \qquad u''(0) < 0. \qquad (2.2)$$

It then follows from a Taylor expansion that the saddlepoint property (1.6) fails: for any $F_1 \in \mathscr{G}_\varepsilon^*(G)$ there is a $t^* = t^*(F_1) \leqslant 1$ such that

$$V(\chi_0, F_t) > V(\chi_0, F_0) \quad \text{for } 0 < t < t^*(F_1). \qquad (2.3)$$

To see that (2.2) holds, first define

$$J(\chi)(x) = 2x\chi'(x) - \chi^2(x), \qquad \dot{S}_0 = \frac{\mathrm{d}}{\mathrm{d}t} S_0(F_t)\bigg|_{t=0}.$$

Lengthy calculations, detailed in Wu (1990), yield

$$u'(0) = \int J(\chi_0)(x) \, \mathrm{d}(F_1 - F_0)(x),$$ (2.4)

$$\dot{S}_0 = \left[\int \chi_0(x) \, \mathrm{d}F_1(x) \right] / I(F_0; 1),$$ (2.5)

$$u''(0) = -\tfrac{1}{2}(\dot{S}_0)^2 \left[\int J^2(\chi_0)(x) \, \mathrm{d}F_0(x) \right.$$

$$\left. + \left\{ \frac{I(F_0; 1)\int \chi_0^4(x) \, \mathrm{d}F_0(x) - \left(\int \chi_0^3(x) \, \mathrm{d}F_0(x) \right)^2}{I(F_0; 1)} \right\} \right].$$ (2.6)

Write $F_i' = f_i = (1 - \varepsilon)g + \varepsilon h_i$, $i = 0, 1$. From (2.1), h_0 has support

$$D(h_0) = \{(-\infty, -x_1) \cup (-x_0, x_0) \cup (x_1, \infty)\}.$$

For $F_1 \in \mathscr{G}_\varepsilon^*(G)$,

$$D(h_1) \subsetneq D(h_0).$$ (2.7)

We have

$$J(\chi_0)(x) = \begin{cases} -k^2, & x \in D(h_0), \\ J(\xi)(x), & x \notin D(h_0). \end{cases}$$ (2.8)

The equality in (2.2) is now immediate from (2.4), (2.7) and (2.8).

Let $j \in \{1, 2\}$ be such that $F_1 \in \mathscr{G}_\varepsilon^j(G)$. Since $S_0(F) = 1$, we have from (2.1), (2.5) and (2.7) that

$$I(F_0; 1)\dot{S}_0 = \int \chi_0(x) \, \mathrm{d}(F_1 - F_0)(x)$$

$$= \varepsilon k \left[(-1)^j + \int_{|x| < x_0} h_0(x) \, \mathrm{d}x - \int_{|x| > x_1} h_0(x) \, \mathrm{d}x \right] \neq 0.$$

Thus $\dot{S}_0 \neq 0$. Since $I(F_0; 1) = \int \chi_0^2(x) \, \mathrm{d}F_0(x)$, the term in braces in (2.6) is non-negative by the Cauchy–Schwarz Inequality. The inequality in (2.2) now follows. This establishes (2.3).

2.2. $\mathscr{G}_\varepsilon(\Phi)$, small ε

If $\mathscr{G} = \Phi$ and $\varepsilon \leqslant \varepsilon_0(\Phi) = 0.2051$, then in (2.1) we have

$$x_0 = 0, \qquad \xi(x) = x^2 - 1, \qquad k = x_1^2 - 1.$$

Define

$$F_- = (1 - \varepsilon)\Phi + \varepsilon\delta_0,$$ (2.9)

where δ_0 is the d.f. with all mass at 0. We claim that (1.6) fails, in that

$$V(\chi_0, F_-) > V(\chi_0, F_0)$$ (2.10)

for $0.0997 \leqslant \varepsilon \leqslant 0.2051$.

Table 1
Asymptotic variances $V(\chi_0, F_0)$ and $V(\chi_0, F_-)$ in $\mathscr{G}_\varepsilon(\Phi)$; F_- given by (2.9), $S_- = S_0(F_-)$

ε	S_-	$V(\chi_0, F_0)$	$V(\chi_0, F_-)$
0.2051	0.5194	1.471	2.775
0.1801	0.5831	1.359	2.081
0.1672	0.6138	1.305	1.829
0.1505	0.6547	1.232	1.553
0.1353	0.6905	1.167	1.356
0.1216	0.7218	1.110	1.210
0.1132	0.7406	1.076	1.132
0.1055	0.7579	1.044	1.067
0.1000	0.7700	1.025	1.029
0.0997	0.7707	1.020	1.021
0.0993	0.7716	1.01844	1.01842

Neither our method of establishing (2.10), nor Huber's (1981) method of showing that (1.6) holds in $\mathscr{G}_\varepsilon(\Phi)$ if $\varepsilon \leqslant 0.04$, yields a sharp bound. We conjecture that there exists $\varepsilon_*(\Phi)$, strictly between 0.04 and 0.0997, with (1.6) holding for $\varepsilon < \varepsilon_*$, failing for $\varepsilon > \varepsilon_*$.

To show (2.10), we first calculate, from (2.9) and (1.3),

$$V(\chi_0, F_0) = \frac{(1-\varepsilon)\left\{\int_{-x_1 S_-}^{x_1 S_-}(x/S_-)^4\,d\Phi(x) + 2x_1^4[1 - \Phi(x_1 S_-)]\right\} - 1}{4\left\{1 - 2(1-\varepsilon)x_1^2[1 - \Phi(x_1 S_-)]\right\}^2}. \tag{2.11}$$

Here, x_1 and S_- are functions of ε through the relationships $\int f_0(x)\,dx = 1$ and $S_- = S_0(F_-)$, i.e.

$$2\left[\Phi(x_1) - \tfrac{1}{2}\right] + \frac{2x_1\phi(x_1)}{x_1^2 - 1} = \frac{1}{1-\varepsilon}, \tag{2.12}$$

$$\int_{-x_2 S_-}^{x_1 S_-}\left(\frac{x}{S_-}\right)^2 d\Phi(x) + 2x_1^2(1 - \Phi(x_1 S_-)) = \frac{1}{1-\varepsilon}. \tag{2.13}$$

Now $V(\chi_0, F_-)$ is a function of ε, through (2.11)–(2.13). Direct numerical calculations give (2.10). See Table 1.

2.3. $\mathscr{X}_\varepsilon(G)$

The minimum information $F_0 \in \mathscr{X}_\varepsilon(G)$ was obtained by Wu (1990). The reader is referred to Wiens and Wu (1990) for a precise listing of the assumptions on G. The most important of these is an assumption concerning the shape of $J(\xi)(x)$, which requires $\xi(x)$ to be an increasing function of $|x|$. These assumptions are satisfied by all of those d.f.'s given in Subsection 2.1.

There are several forms of F_0, depending on the value of ε. In each form, F_0 is symmetric and there are constants $0 < a \leqslant b < c \leqslant d < \infty$ such that:

(a) $G(x) - \varepsilon < F_0(x) < G(x) + \varepsilon$, $|x| \in [0, a) \cup (b, c) \cup (d, \infty)$;

(b) $G(x) = F_0(x) - \varepsilon$, $x \in [a, b]$;

(c) $G(x) = F_0(x) + \varepsilon$, $x \in [c, d]$;

(d) $J(\chi_0)(x)$ is constant on each of $[0, a)$, (b, c) and $[d, \infty)$;

(e) $\chi_0(x)$ is constant on each of $[0, a)$ and $[d, \infty)$, non-constant on (b, c).

Define

$$\mathcal{H}_\varepsilon^*(G) = \{ F_1 \in \mathcal{H}_\varepsilon(G) \mid I(F_1; 1) < \infty, \; F_1(x) \equiv F_0(x)$$
$$\text{whenever } |F_0(x) - G(x)| = \varepsilon, \; S_0(F_1) \neq S_0(F_0) \}.$$

In the presence of the other conditions in the definition of $\mathcal{H}_\varepsilon(G)$, and in view of (e) above, the condition '$S_0(F_1) \neq S_0(F_0)$' is equivalent to

$$\int_b^c \chi_0(x) \, \mathrm{d}(F_1 - F_0)(x) \neq 0.$$

Thus $\mathcal{H}_\varepsilon^*(G) \neq \emptyset$.

We now proceed as in Subsection 2.1. With $u(t)$ as defined there, for $F_1 \in \mathcal{H}_\varepsilon^*(G)$, we again calculate (2.4)–(2.6). That $u'(0) = 0$ follows from the definition of $\mathcal{H}_\varepsilon^*(G)$, and (b)–(d) above. That $u''(0) < 0$ follows exactly as before. Thus, for all ε and each $F_1 \in \mathcal{H}_\varepsilon^*(G)$, there is a $t^*(F_1)$ such that (2.3) holds.

References

Bickel, P.J. and E.L. Lehmann (1976), Descriptive statistics for non-parametric models III. Dispersion, *Ann. Statist.* **4**, 1139–1158.

Boos, D.D. and R.J. Serfling (1980), A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates, *Ann. Statist.* **8**, 618–624.

Collins, J.R. (1983), On the minimax property for R-estimators of location, *Ann. Statist.* **11**, 1129–1195.

Collins, J.R. and D.P. Wiens (1985), Minimax variance M-estimators in ε-contamination models, *Ann. Statist.* **13**, 1078–1096.

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

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

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

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

Sacks, J. and D. Ylvisaker (1982), L- and R-estimation and the minimax property, *Ann. Statist.* **10**, 643–645.

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

Wiens, D. (1986), Minimax variance M-estimators of location in Kolmogorov neighbourhoods, *Ann. Statist.* **13**, 724–732.

Wiens, D. (1987), Robust weighted Cramér–von Mises estimators of location, with minimax variance in ε-contamination neighbourhoods, *Canad. J. Statist.* **15**, 269–278.

Wiens, D.P. and K.H.E. Wu (1990), Distributions minimizing Fisher information for scale in Kolmogorov neighbourhoods, to appear in *Canad. J. Statist.*

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