Jana Jurečková

Locally optimal estimates of location

Commentationes Mathematicae Universitatis Carolinae, Vol. 18 (1977), No. 3, 599--610

Persistent URL: http://dml.cz/dmlcz/105804

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

18,3 (1977)

LOCALLY OPTIMAL ESTIMATES OF LOCATION

Jana JUREČKOVÁ, Praha

Abstract: It is proved that the maximum likelihood estimate is locally optimal estimate of the centre of symmetry of any unimodal symmetric distribution provided its density is absolutely continuous and has integrable derivative. As an application, an L-estimate of the centre of symmetry is suggested which seems to have good local properties with respect to other L-estimates.

Key words: Maximum likelihood estimate, locally most powerful test, locally optimal estimate, L-estimate.

AMS: 62F10, 62G05 Ref. Ž. 9.741

---

1. Introduction. Let $X_1,\dots,X_N$ be independent random variables distributed according to a common density $f(x-\Theta)$, $x \in R^1$, such that $f(x) = f(-x)$, $x \in R^1$. Let $X^{(1)} \leq X^{(2)} \leq \dots \leq X^{(N)}$ be the corresponding order statistics. Assuming that $f(x)$ is absolutely continuous, $\int |f'(x)|\,dx < \infty$ and that f is unimodal, we shall show that the maximum likelihood estimate of Θ is locally optimal in certain sense among all median unbiased estimates of Θ. This property of maximum likelihood estimate is proved with the aid of the theory of locally most powerful tests. In the family of median unbiased L-estimates of Θ, i.e. estimates of the form $\sum_{i=1}^{N} c_i X^{(i)}$, we suggest one which seems to have good local properties in the neighbourhood of Θ.

2. Local optimality of maximum likelihood estimate. Let $X_1,\dots,X_N$ be a random vector distributed according to the density

(2.1) $$p_\Theta(x_1,\dots,x_N) = \prod_{i=1}^{N} f(x_i-\Theta), \quad \Theta \in R^1$$

where f is a known symmetric density. The problem is that of estimating the location parameter Θ by an estimate $\hat{\theta}$ locally optimal in certain sense in a neighbourhood of real value Θ . One of the possible definitions of the local optimality of the estimate $\hat{\Theta}$ is the following

Definition 2.1. We say that the estimate $\hat{\Theta}$ is locally optimal in the set $\mathcal{E}$ of estimates of Θ , if, given any other estimate $\Theta^* \in \mathcal{E}$, there exists an $\varepsilon_0 > 0$ such that

(2.2) $$P_\Theta\{|\hat{\theta} - \Theta| > \varepsilon\} \le P_\Theta\{|\Theta^* - \Theta| > \varepsilon\}$$

holds for all ε, $0 < \varepsilon < \varepsilon_0$ and for all $\Theta \in R^1$.

Theorem 2.1. Let $X_1,\dots,X_N$ be a random vector distributed according to the density (2.1) where f is a known symmetric density, absolutely continuous, unimodal and such that

(2.3) $$\int |f'(x)|dx < \infty \, .$$

Let

(2.4) $$\underline{\Theta} = \inf\{\Theta : -\sum_{i=1}^{N} \frac{f'(X_i-\Theta)}{f(X_i-\Theta)} = 0\},$$

(2.5) $$\overline{\Theta} = \sup\{\Theta : -\sum_{i=1}^{N} \frac{f'(X_i-\Theta)}{f(X_i-\Theta)} = 0\}$$

and

$$(2.6) \qquad \hat{\theta} = \begin{cases} \underline{\theta} & \text{with probability } \frac{1}{2} \\ \overline{\theta} & \text{with probability } \frac{1}{2} \end{cases}$$

where the randomization does not depend on $X_1,\dots,X_N$. Then $\hat{\theta}$ is locally optimal in the set of all median unbiased estimates of θ.

Theorem 2.1 establishes the local optimality of the maximum likelihood estimate. The proof of the theorem will be based on the following lemma.

Lemma 2.1. Under the assumptions of Theorem 2.1, the test with critical function

$$\Phi(x) = 1 \quad \dots \text{if} \quad -\sum_{i=1}^{N} \frac{f'(x_i - \theta_0)}{f(x_i - \theta_0)} > 0$$

$$\Phi(x) = \frac{1}{2} \quad \dots \text{if} \quad -\sum_{i=1}^{N} \frac{f'(x_i - \theta_0)}{f(x_i - \theta_0)} = 0$$

$$\Phi(x) = 0 \quad \dots \text{if} \quad -\sum_{i=1}^{N} \frac{f'(x_i - \theta_0)}{f(x_i - \theta_0)} < 0$$

is locally most powerful in the set of all level $\alpha = \frac{1}{2}$ tests of $H: \theta = \theta_0$ against $K: \theta > \theta_0$.

Proof of Lemma 2.1. Remind that the test $\Phi(x)$ is called locally most powerful test of $\theta = \theta_0$ against $\theta > \theta_0$ on the level α if, given any other α-test Φ^*, there exists an $\varepsilon > 0$ such that $E_\theta \Phi(X) \geq E_\theta \Phi^*(X)$ for all θ, $\theta_0 < \theta < \theta_0 + \varepsilon$.

Assume, without loss of generality, that $\theta_0 = 0$. Let P_θ denote the probability distribution corresponding to p_θ. For any $\theta \neq 0$ and $A \in \mathfrak{B}_N$ we have

$$P_\Theta(A) = \int \underset{A}{\cdots} \int \prod_{i=1}^{N} f(X_i - \Theta)\, dx_1 \ldots dx_N =$$

$$= P_0(A) + \Theta \int \underset{A}{\cdots} \int \frac{1}{\Theta}\Big[\prod_{i=1}^{N} f(x_i - \Theta) - \prod_{i=1}^{N} f(x_i)\Big] dx_1 \ldots$$

(2.8) $\ldots\, dx_N = P_0(A) +$

$$+ \Theta \sum_{i=1}^{N} \int \underset{A}{\cdots} \int \Big[\frac{f(x_i - \Theta) - f(x_i)}{\Theta} \prod_{j=1}^{i-1} f(x_j - \Theta) \prod_{j=i+1}^{N} f(x_j) \, dx_1 \ldots dx_N .$$

It holds

(2.9)
$$\lim_{\Theta \to 0} \Big\{ \frac{1}{\Theta} \big[f(x_i - \Theta) - f(x_i) \big] \prod_{j=1}^{i-1} f(x_j - \Theta) \prod_{j=i+1}^{N} f(x_j) \Big\} = - f'(x_i) \prod_{\substack{j=1 \\ j \neq i}}^{N} f(x_j), \qquad 1 \leq i \leq N,$$

almost everywhere in x. Moreover, if $\Theta > 0$,

(2.10)
$$\int \ldots \int \Big| \frac{1}{\Theta} \big[f(x_i - \Theta) - f(x_i) \big] \Big| \prod_{j=1}^{i-1} f(x_j - \Theta) \prod_{j=i+1}^{N} f(x_j) \, dx_1 \ldots dx_N = \int_{-\infty}^{\infty} \Big| \frac{1}{\Theta} \big(f(x_i - \Theta) - f(x_i) \big) \Big| dx_i =$$
$$= \int_{-\infty}^{\infty} \Big| \frac{1}{\Theta} \int_0^{\Theta} f'(x_i - t) dt \Big| dx_i \leq \frac{1}{\Theta} \int_0^{\Theta} \int_{-\infty}^{\infty} \big| f'(x_i - t) \big| dx_i \, dt,$$

and we get an analogous result if $\Theta < 0$. Thus

(2.11)
$$\limsup_{\Theta \to 0} \int \ldots \int \Big| \frac{1}{\Theta} \big(f(x_i - \Theta) - f(x_i) \big) \Big| \prod_{j=1}^{i-1} f(x_j - \Theta) \prod_{j=i+1}^{N} f(x_j) \, dx_1 \ldots dx_N \leq \int_{-\infty}^{\infty} \big| f'(x_i) \big| dx_i .$$

It then follows from (2.9),(2.11) and from Theorem II.4.2 of [1] that

$$\lim_{\theta\to 0}\sum_{i=1}^{N}\int\underset{A}{\cdots}\int\frac{1}{\theta}\,(f(x_i-\theta)-f(x_i))\prod_{j=1}^{i-1}f(x_j-\theta)$$

(2.12)
$$\prod_{j=i+1}^{N}f(x_j)\,dx_1\ldots dx_N=\sum_{i=1}^{N}\int\underset{A}{\cdots}\int(-f'(x_i))\prod_{\substack{j=1\\ j\neq i}}^{N}f(x_j)\,dx_1\ldots dx_N=\int\underset{A}{\cdots}\int\Big(-\sum_{i=1}^{N}\frac{f'(x_i)}{f(x_i)}\Big)\prod_{j=1}^{N}f(x_j)\,dx_1\ldots dx_N.$$

It follows from (2.8) and (2.12) that

(2.13)
$$\lim_{\theta\to 0}\frac{1}{\theta}\left[P_\theta(A)-P_0(A)\right]=\int\underset{A}{\cdots}\int\Big(-\sum_{i=1}^{N}\frac{f'(x_i)}{f(x_i)}\Big)\prod_{j=1}^{N}f(x_j)dx_1\ldots dx_N$$

holds for any $A\in\mathcal{B}_N$.

Let us denote

(2.14)
$$S(x-\theta)=-\sum_{i=1}^{N}\frac{f'(x_i-\theta)}{f(x_i-\theta)}$$

and

(2.15) $A=\{x:\ S(x)>0\}$; $A'=\{x:\ S(x)\geq 0\}$.

Let Φ be the test defined in (2.7) and let Φ^* be any other level $\frac{1}{2}$ test; denote

(2.16) $B=\{x:\ \Phi^*(x)=1\}$, $B'=\{x:\ \Phi^*(x)>0\}$.

Then (2.13), (2.15) and (2.16) imply

(2.17)
$$\lim_{\theta\to 0}\frac{1}{\theta}\left[E_\theta\Phi(X)-E_0\Phi(X)\right]=$$

$$=\lim_{\theta\to 0}\frac{1}{2\theta}\left[P_\theta(A)+P_\theta(A')-P_0(A)-P_0(A')\right]=$$

$$=\int\underset{\{x:\,S(x)>0\}}{\cdots}\int S(x)\prod_{j=1}^{N}f(x_j)\,dx_1\ldots dx_N$$

and

$$(2.18)\qquad \lim_{\theta\to 0} \frac{1}{\theta}[E_\theta \Phi^*(\mathbf{X}) - E_0\Phi^*(\mathbf{X})] \leq$$

$$\leq \int \underset{B'}{\cdots} \int S(\mathbf{x}) \prod_{j=1}^{N} f(x_j)\, dx_1 \ldots dx_N \leq$$

$$\leq \int \underset{S(\mathbf{x})>0}{\cdots} \int S(\mathbf{x}) \prod_{j=1}^{N} f(x_j)\, dx_1 \ldots dx_N .$$

(2.17) and (2.18) mean that the function $E_\theta(\Phi(\mathbf{X}) - \Phi^*(\mathbf{X}))$ is nondecreasing at $\theta = 0$, so that there is an $\varepsilon > 0$ such that this function is nonnegative for $0 < \theta < \varepsilon$.

<u>Proof of Theorem 2.1</u>. The symmetry and the unimodality of f imply that $S(\mathbf{x}-\theta)$ is nonincreasing in θ for any fixed $\mathbf{x}$ and that

$$(2.19)\qquad P_\theta(S(\mathbf{X}-\theta)< 0) = P_\theta(S(\mathbf{X}-\theta)>0) \leq \frac{1}{2}.$$

Moreover, the function $-\sum_{i=1}^{N} \frac{f'(x_i)}{f(x_i)}$ is, as a finite sum of nondecreasing functions, continuous almost everywhere in $\mathbf{x} = (x_1, \ldots, x_N)$ and thus $S(\mathbf{x}-\theta)$ is continuous in θ for almost all θ and almost all $\mathbf{x}$.

The set $\{\theta : S(\mathbf{x}-\theta) \leq 0\}$ is a half-line; denote

$$(2.20)\qquad \underline{\theta}(\mathbf{x}) = \inf\{\theta : S(\mathbf{x}-\theta) \leq 0\} .$$

It follows from the continuity mentioned above that

$$(2.21)\qquad P_\theta(S(\mathbf{X}-\underline{\theta}) = 0) = 1 \text{ for any } \theta \in R^1.$$

Analogously, put

$$(2.22)\qquad \overline{\theta}(\mathbf{x}) = \sup\{\theta : S(\mathbf{x}-\theta) \geq 0\}$$

and

$$(2.23) \qquad \hat{\Theta}(x) = \begin{cases} \underline{\Theta}(x) & \text{with probability } \frac{1}{2} \\ \overline{\Theta}(x) & \text{with probability } \frac{1}{2} \end{cases}$$

where the randomization does not depend on X. Then $\hat{\Theta}$ is median unbiased; actually,

$$(2.24) \qquad P_\Theta(\hat{\Theta} < \Theta) = \frac{1}{2}\left[P_\Theta(\underline{\Theta} < \Theta) + P_\Theta(\overline{\Theta} < \Theta)\right] =$$

$$= \frac{1}{2} P_0(S(X) \leq 0) + \frac{1}{2} P_0(S(X) < 0) = \frac{1}{2}$$

and we get $P_\Theta(\hat{\Theta} > \Theta) = \frac{1}{2}$ analogously. Let Θ^* be any other median unbiased estimate. Then the sets $\{x: \Theta^*(x) < \Theta_0\}$ and $\{x: \hat{\Theta}(x) < \Theta_0\}$ can be considered as acceptance regions of level $\frac{1}{2}$ tests of $H_{\Theta_0}: \Theta = \Theta_0$ against $K_{\Theta_0}: \Theta > \Theta_0$ for any Θ_0; the second of them being locally most powerful in view of Lemma 2.1. Consequently, there is an ε_0 such that the first test is dominated by the second one for $\Theta_0 < \Theta < \Theta_0 + \varepsilon_0$. Let us fix an ε, $0 < \varepsilon < \varepsilon_0$. Then it holds

$$(2.25) \qquad P_{\Theta_1}(\hat{\Theta} < \Theta_1 - \varepsilon) \leq P_{\Theta_1}(\Theta^* < \Theta_1 - \varepsilon) \text{ for any } \Theta_1 \in R^1.$$

Actually, $\{x: \hat{\Theta} < \Theta_1 - \varepsilon\}$ and $\{x: \Theta^* < \Theta_1 - \varepsilon\}$ can be considered as acceptance regions of level $\frac{1}{2}$ tests of the hypothesis $H: \Theta = \Theta_1 - \varepsilon (= \Theta_0)$ against $K: \Theta > \Theta_1 - \varepsilon$; the first one being locally most powerful. (2.25) then follows from Lemma 2.1. Analogously, we derive

$$(2.26) \qquad P_{\Theta_1}(\hat{\Theta} > \Theta_1 + \varepsilon) \leq P_{\Theta_1}(\Theta^* > \Theta_1 + \varepsilon)$$

for all $\Theta_1 \in R^1$ and $0 < \varepsilon < \varepsilon_0$.

3. Application: an L-estimate of location with good local properties. We have shown that the estimate $\hat{\theta}$ defined in (2.4) - (2.6) is locally optimal in the sense of Definition 2.1 among all median unbiased estimates of θ. It follows from Theorem 2.1 that $\hat{\theta}$ is locally optimal also among all median unbiased estimates based on the order statistics $X^{(1)} \leq \ldots \leq X^{(N)}$. Suppose now that we wish to estimate θ by a median unbiased L-estimate, i.e. by an estimate of the form

$$(3.1) \qquad \tilde{\theta} = \sum_{i=1}^{N} c_i X^{(i)}$$

which has good local properties relative to other median unbiased L-estimates. Such estimate must be a good approximation of $\hat{\theta}$ with respect to other L-estimates; both estimates may coincide for some special distributions.

The problem of the best approximation of $\hat{\theta}$ by an L-estimate has been solved asymptotically as $N \to \infty$: suppose that the coefficients c_i in (3.1) are generated by a function $J(t)$, $J(1-t) = J(t)$, $0 < t < 1$, in the following way:

$$(3.2) \qquad c_i = \frac{1}{N} J\left(\frac{i}{N+1}\right), \quad i = 1,\ldots,N.$$

If $N \to \infty$ and various regularity conditions are satisfied, then the function

$$(3.3) \quad J(t) = \frac{d\varphi(t,f)}{dt} \cdot f(F^{-1}(t)) \left[\int_0^1 \frac{d\varphi(t,f)}{dt} f(F^{-1}(t))dt\right]^{-1}$$

where

$$(3.4) \qquad \varphi(t,f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \quad 0 < t < 1,$$

yields an asymptotically efficient estimate, i.e. one which achieves the information inequality lower bound as $N \longrightarrow \infty$ (Jung [2]).

The problem of the best approximation of $\hat{\theta}$ by an L-estimate has not yet been solved from the local point of view. Here we shall only suggest a member of the family of L-estimates which seems to have good local properties.

Suppose that $\varphi(t,f)$ given in (3.4) is continuous at $t = \frac{1}{N-1}, \ldots, \frac{N-2}{N-1}$ and that

$$\lim_{t \to 0} \frac{\varphi(t,f)}{F^{-1}(t)} < \infty \; ; \tag{3.5}$$

moreover, let N = 2n. Put

$$c_1 = c_N = \frac{1}{K} \lim_{t \to 0} \frac{\varphi(t,f)}{F^{-1}(t)} \tag{3.6}$$

and

$$c_i = c_{N-i+1} = \frac{1}{K} \cdot \frac{1}{F^{-1}\left(\frac{i-1}{N-1}\right)} \cdot \varphi\left(\frac{i-1}{N-1}, f\right), \; i=2,\ldots,n, \tag{3.7}$$

where

$$K = 2 \lim_{t \to 0} \frac{\varphi(t,f)}{F^{-1}(t)} + 2 \sum_{i=2}^{m} \frac{1}{F^{-1}\left(\frac{i-1}{N-1}\right)} \cdot \varphi\left(\frac{i-1}{N-1}, f\right). \tag{3.8}$$

The coefficients c_i given by (3.6) and (3.7) imply relatively small values of the sums

$$\sum_{i=1}^{N} P_\theta \left\{ \left| - \frac{f'(X^{(i)} - \theta)}{f(X^{(i)} - \theta)} - K c_i \, (X^{(i)} - \theta) \right| \geq \varepsilon \right\} \tag{3.9}$$

for sufficiently small ε, $0 < \varepsilon < \varepsilon_0$. Actually, we may write

$$P_\theta\left\{\left|\frac{f'(X^{(i)}-\theta)}{f(X^{(i)}-\theta)} - K c_i (X^{(i)}-\theta)\right| \ge \varepsilon\right\} =$$

$$(3.10) \quad = P_0\left\{\left|-\frac{f'(X^{(i)})}{f(X^{(i)})} - K c_i X^{(i)}\right| \ge \varepsilon\right\} =$$

$$= N \binom{N-1}{i-1} \int_{A_i} t^{i-1}(1-t)^{N-i}\, dt$$

where

$$(3.11) \quad A_i = \left\{t \in (0,1):\ \left|-\frac{f'(F^{-1}(t))}{f(F^{-1}(t))} - K c_i F^{-1}(t)\right| \ge \varepsilon\right\},$$

$$i = 1,\dots,n.$$

The density of beta distribution which appears in the last integral is unimodal with the unique mode at $\hat{t}_i = \frac{i-1}{N-1}$, $i = 1,\dots,n$. Considering a fixed i, $2 \le i \le n$, we see that c_i given by (3.7) eliminates the expression

$$\left|\frac{f'(F^{-1}(t))}{f(F^{-1}(t))} - K c_i F^{-1}(t)\right|$$

just at $t = \hat{t}_i$ and thus to any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$(3.12) \quad \left|t-\hat{t}_i\right| < \delta \Longrightarrow \left|-\frac{f'(F^{-1}(t))}{f(F^{-1}(t))} - K c_i F^{-1}(t)\right| < \varepsilon .$$

It implies that

$$(3.13) \quad N\binom{N-1}{i-1} \int_{A_i} t^{i-1}(1-t)^{N-i} dt =$$

$$= N\binom{N-1}{i-1} \int_{t\in A_i, |t-t_i| \ge \delta} t^{i-1}(1-t)^{N-i}\, dt$$

so that the interval around $\hat{t}_i$ with the highest values of the integrand does not belong to the integration domain.

Similar considerations could be made for $i = 1$. On the other hand, considering any other choice of c_i's, there is an i, $1 \leq i \leq n$ and $\delta > 0$ such that $|t-\hat{t}_i| < \delta$ implies

$$\left| - \frac{f'(F^{-1}(t))}{f(F^{-1}(t))} - K\, c_i\, F^{-1}(t) \right| \geq \varepsilon \quad \text{for sufficiently}$$

small $\varepsilon > 0$. A neighbourhood of $\hat{t}_i$ with the highest values of the integrand is thus a part of the integration domain in (3.10).

As an illustration, consider three most typical unimodal symmetric density shapes.

(i) $$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \qquad x \in R^1$$

(standard normal distribution)

$c_i = \frac{1}{N}$, $i = 1,\dots,N$

$\tilde{\theta} = \bar{X} = \hat{\theta}$.

(ii) $$f(x) = \frac{1}{2}\, e^{-|x|}\,, \qquad x \in R^1$$

(double exponential distribution)

$c_o = c_N = 0$

$$c_i = c_{N-i+1} = \left[K \log \frac{N-1}{2(i-1)}\right]^{-1} \qquad i = 2,\dots,n$$

$$K = 2 \sum_{i=2}^{m} \left[\log \frac{N-1}{2(i-1)}\right]^{-1}$$

$$c_n = \left[K \log \frac{N-1}{N-2}\right]^{-1}$$

(iii) $$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}\,, \qquad x \in R^1$$

(logistic distribution)

$$c_o = c_N = 0$$

$$c_i = c_{N-i+1} = 2\left(\frac{N+1}{2} - i\right)\left(K \log \frac{N-i}{i-1}\right)^{-1}, \quad i = 1,2,\ldots n.$$

$$K = 4 \sum_{i=2}^{n} \left(\frac{N+1}{2} - i\right)\left(\log \frac{N-i}{i-1}\right)^{-1}$$

$$c_n = \left(K \log \frac{N}{N-2}\right)^{-1}.$$

R e f e r e n c e s

[1] J. HÁJEK, Z. ŠIDÁK: Theory of rank tests, Academia, Praha, 1967.

[2] J. JUNG: On linear estimates defined by a continuous weight function, Ark. Math. 3(1955), 199-209.

[3] E.L. LEHMANN: Testing statistical hypotheses, Wiley, 1959.

Matematicko-fyzikální fakulta

Universita Karlova

Sokolovská 83, 18600 Praha 8

Československo

(Oblatum 6.6. 1977)