

ON LOCATION AND SCALE MAXIMUM LIKELIHOOD ESTIMATORS

A. K. GUPTA AND G. J. SZÉKELY

(Communicated by Wei Y. Loh)

ABSTRACT. The maximum likelihood estimators of location and scale parameters have the following undesirable properties: (i) they are not always consistent, and (ii) the asymptotic correlation of the MLE's of the "unrelated" location and scale parameters can be arbitrarily close to one.

In this paper methods to overcome these undesirable properties have been studied.

1. CONSISTENCY OF LOCATION-SCALE TRUNCATED MLE'S

It is well known that the maximum likelihood estimators (MLE's) of the location and scale parameters may not exist. An example by Pitman [7, Chapter 8] is the following. If $h(x) = c(x \log^2 x)^{-1}$ (c is a suitable constant) then the MLE's for the location and scale parameters for the probability density function $\sigma^{-1}h((x - \mu)/\sigma)$ do not exist. On the other hand if h is bounded, upper semicontinuous and $|x|^{1+\alpha}h(x)$ is bounded for some $\alpha > 0$ then the MLE of (μ, σ) is (strongly) consistent. Using some sophisticated results from category theory we could also construct a bounded, noncontinuous probability density function h such that the MLE of the location parameter μ of $h(x - \mu)$ exists but it is not consistent. (In this paper we do not want to discuss the details of the construction of h , we only mention that it can be a mixture of a normal density and a uniform density on a tricky unbounded set.) It is clear that without some continuity conditions we cannot even guarantee the mere existence of MLE. On the other hand we shall see that a simple modification of MLE is always consistent for all continuous strictly positive densities.

Replace by C (> 0) every sample element bigger than C and by $-C$ every sample element smaller than $-C$. Denote by $h_{\mu, \sigma}^*$ the density function of this truncated sample (with respect to the corresponding measure ν).

Received by the editors May 26, 1992.

1991 *Mathematics Subject Classification.* Primary 62E10; Secondary 62H20.

Key words and phrases. Consistency, maximum likelihood, uncorrelated (orthogonal) estimators, truncated estimators.

The first author's research was partially supported by a FRC Major Grant, Bowling Green State University. The second author's research was supported by the Hungarian National Foundation for Scientific Research No. 1405 and 1905.

Theorem 1. *If h is continuous (or at least upper semicontinuous) and strictly positive everywhere, $(\mu, \sigma) \in \Theta: = [a, b] \times [c, d]$ where $b > a$, $d > c > 0$ are arbitrary (finite) real numbers then the MLE based on $h_{\mu, \sigma}^*$ is always (strongly) consistent. (If $c = 0$ were not excluded then the MLE might even not exist as the above mentioned example by Pitman shows.)*

Proof. We are going to prove somewhat more than strong consistency. Let $(\mu_0, \sigma_0) \in \Theta$ be the true parameter. If $\Theta_1 \subset \Theta$ is any closed set and $(\mu_0, \sigma_0) \notin \Theta_1$ then

$$\int \left(\log \frac{h_{\mu, \sigma}^*}{h_{\mu_0, \sigma_0}^*} \right) h_{\mu_0, \sigma_0}^* d\nu$$

is a continuous function of (μ, σ) and (by Jensen's inequality) its maximum on Θ_1 is less than 0. By the strong law of large numbers and the above mentioned continuity in (μ, σ) we have

$$\lim_{n \rightarrow \infty} \sup_{(\mu, \sigma) \in \Theta_1} \frac{1}{n} \sum_{i=1}^n \log \frac{h_{\mu, \sigma}^*(x_i)}{h_{\mu_0, \sigma_0}^*(x_i)} < 0$$

with probability one (x_1, x_2, \dots, x_n denote the sample elements); thus,

$$\sup_{(\mu, \sigma) \in \Theta_1} \prod_{i=1}^n h_{\mu, \sigma}^*(x_i) < \prod_{i=1}^n h_{\mu_0, \sigma_0}^*(x_i)$$

if n is large enough (with probability one). Thus the truncated MLE is strongly consistent. (In fact, we proved somewhat more than required, since strong consistency already follows if Θ_1 is any one-point set.)

2. ON THE DEPENDENCE OF LOCATION AND SCALE MAXIMUM LIKELIHOOD ESTIMATORS

Denote by $\hat{\Theta}$ the range of MLE's $(\hat{\mu}, \hat{\sigma})$ of (μ, σ) . It is clear that no matter what the density function h is, the set $\hat{\Theta}$ is always the direct product of the range of $\hat{\mu}$ and the range of $\hat{\sigma}$ (observe that a shift of the sample shifts $\hat{\mu}$). Thus no information on either parameter can influence the range of the other one. In spite of this "unrelatedness" of $\hat{\mu}$ and $\hat{\sigma}$ we shall see that their stochastic dependence can be very strong.

Under certain regularity conditions (the support of h is the whole real line, h is three times continuously differentiable, $\int_{-\infty}^{+\infty} (h'(x))^2 x^i / h(x) dx$ is finite for $i = 0, 1, 2$, etc. (see [5, Theorem 6.4.1]) we have

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow N(0, \Sigma) \quad \text{as } n \rightarrow \infty$$

where $\Sigma^{-1} = I(\theta)$ is the Fisher information matrix:

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} I_0 & I_1 \\ I_1 & I_2 - 1 \end{bmatrix},$$

$I_i = \int_{-\infty}^{+\infty} (h'(x)x^i/h(x)) dx$, $i = 0, 1, 2$. This implies that if h is symmetric around 0 then $I_1 = 0$. Thus $\hat{\mu}$, $\hat{\sigma}$ are asymptotically uncorrelated (and asymptotically jointly normal and, hence, asymptotically independent). In general we have that the squared correlation

$$\text{corr}^2(\hat{\mu}, \hat{\sigma}) \sim \frac{I_1^2}{I_0(I_2 - 1)}$$

and the Cauchy-Schwartz inequality shows that

$$\begin{aligned} I_1^2 &= \left[\int \left(\frac{h'(x)}{\sqrt{h(x)}} x - \sqrt{h(x)} \right) \frac{h'(x)}{\sqrt{h(x)}} dx \right]^2 \\ &\leq \int \left(\frac{h'(x)}{\sqrt{h(x)}} x - \sqrt{h(x)} \right)^2 dx \int \frac{h'(x)}{\sqrt{h(x)}} dx = (I_2 - 1)I_0 \end{aligned}$$

with equality iff $h'(x)(x - c) = h(x)$ where c is some constant. Thus the absolute value of this correlation is 1 iff h is linear (on the whole real line), and this is clearly impossible. In the next section we are going to show that $\text{corr}(\hat{\mu}, \hat{\sigma})$ can be arbitrarily close to one.

3. HIGH ASYMPTOTIC CORRELATION OF LOCATION AND SCALE MLE'S

Theorem 2. For every $\varepsilon > 0$ and sufficiently large n there exists a location-scale (μ, σ) -family of distributions such that for the MLE's $\hat{\mu}_n, \hat{\sigma}_n$ we have

$$|\text{corr}(\hat{\mu}_n, \hat{\sigma}_n)| > 1 - \varepsilon.$$

Proof. Take any smooth density function h which satisfies the above mentioned regularity conditions of Lehmann Theorem 6.4.1 and $h'(0) > 0$. Insert a linear segment into this density in the following way:

$$h_\varepsilon(x) = \begin{cases} c_\varepsilon h(x) & \text{for } x \geq 0, \\ c_\varepsilon(ax + b) & \text{for } -(1 - \varepsilon)b/a \leq x < 0, \\ c_\varepsilon \varepsilon h(\frac{1}{\varepsilon}(x + (1 - \varepsilon)b/a)) & \text{for } x < -(1 - \varepsilon)b/a \end{cases}$$

where $a = h'(0)$, $b = h(0)$, and $c_\varepsilon > 0$ is a suitable constant that makes h_ε a density function. We can see that

$$c_\varepsilon \rightarrow \left[\frac{b^2}{2a} + \int_0^\infty h(x) dx \right]^{-1} = c_0 \quad \text{as } \varepsilon \rightarrow 0$$

and

$$h'_\varepsilon(x) = \begin{cases} c_\varepsilon h'(x) & \text{for } x \geq 0, \\ c_\varepsilon a & \text{for } -(1 - \varepsilon)b/a \leq x < 0, \\ c_\varepsilon h'(\frac{1}{\varepsilon}(x + (1 - \varepsilon)b/a)) & \text{for } x < -(1 - \varepsilon)b/a. \end{cases}$$

Now

$$I_i = \int_{-\infty}^{+\infty} \frac{h'_\varepsilon(x)^2}{h_\varepsilon(x)} x^i dx = \left(\int_0^\infty + \int_{-(1-\varepsilon)b/a}^0 + \int_{-\infty}^{-(1-\varepsilon)b/a} \right) \frac{h'_\varepsilon(x)^2}{h_\varepsilon(x)} x^i dx.$$

Here the second integral $\rightarrow \infty$ as $\varepsilon \rightarrow 0$, while the first and third integrals are bounded (thus negligible). To see this observe that

$$c_\varepsilon \int_0^\infty \frac{h'(x)^2}{h(x)} x^i dx = O(1).$$

Now introduce $y = \frac{1}{\varepsilon}(x + (1 - \varepsilon)b/a)$ and

$$c_\varepsilon \int_{-\infty}^0 \left(\varepsilon y - \frac{(1 - a)b}{a} \right)^i \frac{h'(y)}{h(y)} dy = O(1).$$

Further let $z = ax + b$; then

$$c_\varepsilon \int_{-(1-\varepsilon)b/a}^0 \frac{a^2 x^i}{ax+b} dx = c_\varepsilon \int_{\varepsilon b}^b \left(\frac{z-b}{a}\right)^i \frac{adz}{z} = c_\varepsilon a \left(-\frac{b}{a}\right)^i \log \frac{1}{\varepsilon} + O(1).$$

Thus, $I_i \sim c_0 a (-b/a)^i \log 1/\varepsilon$, and as $\varepsilon \rightarrow 0$,

$$\text{corr}^2(\hat{\mu}_n, \hat{\sigma}_n) \sim \frac{I_1^2}{I_0(I_2 - 1)} \sim 1.$$

Finally, it may be noted that the regularity conditions guarantee (see, e.g., [1, Chapter 2, Theorem 25.1]) the existence of arbitrarily many moments of $\hat{\mu}_n$ and $\hat{\sigma}_n$ and thus the existence of $\text{Corr}(\hat{\mu}_n, \hat{\sigma}_n)$.

Remark 1. High correlation indicates that the ratio of the eigenvalues of the covariance matrix of $(\hat{\mu}, \hat{\sigma})$ can be large. In fact, we can prove that this ratio tends to ∞ as $\varepsilon \rightarrow 0$. This is clearly an undesirable property (it shows a kind of inaccuracy of the estimators). A linear reparametrization can always make the MLE's $\hat{\mu}_n$ and $\hat{\sigma}_n$ uncorrelated. If Σ_n is the covariance matrix of $(\hat{\mu}_n, \hat{\sigma}_n)$ then $\tilde{\Sigma}_n = \sigma^2 \Sigma_n$ does not depend on (μ, σ) . Thus if

$$\tilde{\Sigma}_n^{-1} = \begin{pmatrix} c_{0n} & c_{1n} \\ c_{1n} & c_{2n} \end{pmatrix},$$

then an easy computation shows that the linear transforms $\mu^* = \mu + \sigma c_{1n}/c_{0n}$, $\sigma^* = \sigma$ have uncorrelated MLE's. (Observe that $c_{1n}/c_{0n} \rightarrow I_1/I_0$ as $n \rightarrow \infty$.) This kind of orthogonal reparametrization is not always easy, sometimes not even possible (see, e.g., [3, 4, 6, 2]).

Remark 2. In the regular case (when the MLE's are asymptotically jointly normal) uncorrelatedness implies asymptotic independence. A typical nonregular example is the uniform distribution on some interval (a, b) . Here the MLE's of location, $\mu = (a+b)/2$, and scale $\sigma = b-a$ parameters are $\hat{\mu} = (\hat{a} + \hat{b})/2$ and $\hat{\sigma} = \hat{b} - \hat{a}$ respectively, where $\hat{a} = \min X_i$ and $\hat{b} = \max X_i$. One can easily show that $\text{corr}(\hat{\mu}, \hat{\sigma}) = 0$, but $\hat{\mu}, \hat{\sigma}$ are not independent, not even asymptotically. To see this observe that $n(\hat{a} - a)$ and $n(b - \hat{b})$ are asymptotically independent and exponentially distributed. Thus their sum and their difference (hence $\hat{\mu}$ and $\hat{\sigma}$) cannot be independent by a theorem of Bernstein (see, e.g., [8]), which claims that for independent random variables X and Y the sum $X + Y$ and the difference $X - Y$ are independent iff X and Y are normally distributed.

REFERENCES

1. A. A. Borokov, *Mathematical statistics*, Nauka, Moscow, 1984.
2. D. R. Cox and N. Reid, *Parameter orthogonality and approximate condition inference (with discussion)*, J. Roy. Statist. Soc. B **49** (1987), 1-39.
3. V. S. Huzurbazar, *Probability distribution and orthogonal parameters*, Proc. Cambridge Philos. Soc. **46** (1950), 281-284.
4. ———, *Sufficient statistics and orthogonal parameters*, Sankhyā **17** (1956), 217-220.
5. E. L. Lehmann, *Theory of point estimation*, Wiley, New York, 1983.
6. A. F. S. Mitchell, *Sufficient statistics and orthogonal parameters*, Proc. Cambridge Philos. Soc. **58** (1962), 326-337.

7. E. J. G. Pitman, *Some basic theory of statistical inference*, Chapman and Hall, London, and Wiley, New York, 1979.
8. A. Rényi, *Probability theory*, North-Holland, Amsterdam, 1970.

DEPARTMENT OF MATHEMATICS & STATISTICS, BOWLING GREEN STATE UNIVERSITY, BOWLING GREEN, OHIO 43403

E-mail address: gupta@andy.bgsu.edu

DEPARTMENT OF MATHEMATICS, EÖTVÖS LORAND UNIVERSITY, EGYETEM-TER 1-3, 1364 BUDAPEST, HUNGARY