KYBERNETIKA — VOLUME 31 (1995), NUMBER 3, PAGES 297–302

FINITE SAMPLE OPTIMALITY OF EXTREMES AND MID-RANGE

CHRIS A. J. KLAASSEN[1]

Consider estimation of the location of the center of a uniform distribution by a translation equivariant estimator. It is shown that for a broad class of loss functions and for each sample size the sample mid-range is an optimal estimator. Similarly, the sample minimum is optimal in estimating the start of a nonincreasing density.

1. HÁJEK AND EFFICIENT ESTIMATION

This paper is the outgrowth of a lecture held at the Minisymposium in Honour of Jaroslav Hájek, June 8, 1994. It was entitled "Hájek and Efficient Estimation" and it discussed shortly two papers of Hájek's, which are great milestones in the development of the asymptotic theory of efficient estimation. Of course, these are the famous papers on the Hájek–Le Cam convolution theorem and Local Asymptotic Minimax theorem; see Hájek [2, 3]. In fact, these results formulate asymptotic bounds on the performance of estimators and determine the limit behavior of asymptotically efficient estimators. A nice paper discussing these results as consequences of Le Cam's theory on "Limits of Experiments" is van der Vaart [9].

In both papers of Hájek the argument is based crucially on the assumption of Local Asymptotic Normality. If one has n observations stemming from i.i.d. random variables then LAN occurs if one has a regular parametric model. Then good estimators estimate the parameter at rate $1/\sqrt{n}$. In nonregular i.i.d. models, often LAN does not hold and a much faster rate like $1/n$, is possible, typically. So, for applications the regular case is by far the most important one, since in nonregular cases one may estimate the parameter at a better rate anyhow. Moreover, the predominant case is the regular one.

Nevertheless, we will discuss a couple of very specific nonregular situations in this paper, two of which are parametric and one semiparametric. As we will see it is possible there to formulate bounds on the performance of estimators, and to construct estimators attaining these bounds. It should be stressed that these bounds

[1] Part of this research was done at the Department of Probability and Mathematical Statistics of the Charles University at Prague, while the author was visiting in the framework of the exchange program of the Charles University with the University of Amsterdam.

are valid for finite sample sizes, that the proofs are extremely simple, but that they apply only in a few specific models. Therefore, they might contribute to a proper appreciation of the profoundness of the convolution and LAM theorem.

2. AN ELEMENTARY INEQUALITY

Let $X_1, \dots, X_n$ be i.i.d. random variables with density $f(\cdot - \theta)$ with respect to Lebesgue measure on $\boldsymbol{R}$, and with $\theta \in \boldsymbol{R}$. In this classical location problem we will consider the natural class of translation equivariant estimators. The estimator $T_n = t_n(X_1, \dots, X_n)$ is called translation equivariant if the law of $T_n - \theta$ under θ is the same for all θ; or more strictly, if

$$t_n(x_1 + a, \dots, x_n + a) = t_n(x_1, \dots, x_n) + a \tag{2.1}$$

for all $x_1, \dots, x_n,\ a \in \boldsymbol{R}$. Many papers have been written on this topic, most of them considering the regular case essentially, in which $f(\cdot)$ has finite Fisher information for location $\int (f'/f)^2 f$. To mention, but a few: Stone [8], Klaassen [4].

Here we will consider some irregular cases, in which the density $f(\cdot)$ has a jump at the boundary of its support. Our approach will be based on an elementary inequality which involves two functions. One is the distribution function $G_n(\cdot)$ of the normed estimator, more precisely,

$$G_n(y) = P_{f(\cdot - \theta)}(n(T_n - \theta) \le y) = P_f(nT_n \le y), \quad y \in \boldsymbol{R}. \tag{2.2}$$

The other depends only on the density $f(\cdot)$ of the observations, and may take the value ∞ ;

$$c(\eta) = \sup\left\{ \frac{f(x - \eta)}{f(x)} \mid f(x) > 0 \right\}, \quad \eta \in \boldsymbol{R}. \tag{2.3}$$

Clearly, we have for all $y, \theta \in \boldsymbol{R}$,

$$\begin{aligned} 1 - G_n(y - \theta) &= P_f(nT_n > y - \theta) = P_f\left(nt_n\left(X_1 + \frac{\theta}{n}, \dots, X_n + \frac{\theta}{n}\right) > y\right) \\ &= \int \cdots \int \mathbf{1}_{[nt_n(x_1, \dots, x_n) > y]} \prod_{i=1}^n \frac{f(x_i - \frac{\theta}{n})}{f(x_i)} \prod_{i=1}^n f(x_i)\, \mathrm{d}x_i \\ &+ \int \cdots \int \mathbf{1}_{[nt_n(x_1, \dots, x_n) > y]} \mathbf{1}_{[\prod_{i=1}^n f(x_i) = 0]} \prod_{i=1}^n f\left(x_i - \frac{\theta}{n}\right) \mathrm{d}x_i \\ &\le c^n\left(\frac{\theta}{n}\right)(1 - G_n(y)) + P_f\left(\prod_{i=1}^n f\left(X_i + \frac{\theta}{n}\right) = 0\right). \end{aligned} \tag{2.4}$$

As we will see in the subsequent sections, this simple inequality may be used to derive optimality properties of the mid-range and sample extremes. Therefore, we state this inequality and an analogue of it explicitly in a

Proposition 2.1. Under condition (2.1) and notation (2.2), (2.3) the following inequalities hold

$$1 - G_n(y-\theta) \le c^n\left(\frac{\theta}{n}\right)(1 - G_n(y)) + P_f\left(\prod_{i=1}^n f\left(X_i + \frac{\theta}{n}\right) = 0\right), \quad y, \theta \in \boldsymbol{R}, \tag{2.5}$$

$$G_n(y+\theta) \le c^n\left(-\frac{\theta}{n}\right) G_n(y) + P_f\left(\prod_{i=1}^n f\left(X_i - \frac{\theta}{n}\right) = 0\right), \quad y, \theta \in \boldsymbol{R}. \tag{2.6}$$

3. FINITE SAMPLE EFFICIENCY OF SAMPLE MID-RANGE

Consider estimation of the midpoint θ of arbitrary uniform distributions. In the notation of the preceding section this means that we parametrize their densities by

$$f_\zeta(x-\theta) = \frac{1}{2\zeta}\mathbf{1}_{(-\zeta,\zeta)}(x-\theta), \quad \theta \in \boldsymbol{R},\ \zeta > 0. \tag{3.1}$$

Fix ζ, call $f_\zeta(\cdot) = f(\cdot)$ and note that $c(\eta) = \mathbf{1}_{(-2\zeta,2\zeta)}(\eta)$. Consequently, applying (2.5) with $0 < 2y = \theta < 2\zeta n$ we obtain

$$G_n(y) - G_n(-y) \le 1 - \left(1 - \frac{y}{\zeta n}\right)^n, \quad y > 0. \tag{3.2}$$

We will say that the distribution F is at least as concentrated symmetrically as the distribution G if

$$G(y) - G(-y) \le F(y) - F(-y), \quad y > 0, \tag{3.3}$$

and we will denote this ordering by

$$G \ge_s F. \tag{3.4}$$

With

$$F_n(y) = \begin{cases} \frac{1}{2}\left(1 + \frac{y}{\zeta n}\right)^n, & y \le 0, \\ 1 - \frac{1}{2}\left(1 - \frac{y}{\zeta n}\right)^n, & y > 0, \end{cases} \tag{3.5}$$

inequality (3.2) just states

$$G_n \ge_s F_n. \tag{3.6}$$

A maximum likelihood estimator $\hat{\theta}_n$ for θ is the mid-range $\hat{\theta}_n = \frac{1}{2}\left(X_{(1)} + X_{(n)}\right)$; here $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ are the order statistics of $X_1, \ldots X_n$. Some computation shows that $n(\hat{\theta}_n - \theta)$ under θ has distribution function F_n from (3.5). Consequently, the mid-range is optimal in the sense of (3.2) and (3.6), and these inequalities are sharp. In other words, the distribution of the mid-range is at least as concentrated symmetrically as any other translation equivariant estimator of the midpoint of a uniform distribution.

The asymptotic behavior of the mid-range under symmetric densities with bounded support is treated in Bingham [1].

To conclude this section we interpret (3.4) in terms of risks for the classical class $\mathcal{L}_s$ of symmetric loss functions $\ell(\cdot)$; $\ell : \boldsymbol{R} \to [0,\infty)$ belongs to $\mathcal{L}_s$ if $\ell(0) = 0$, $\ell(x) = \ell(|x|)$ and $\ell(\cdot)$ is nondecreasing on $[0,\infty)$. Since for each distribution function F and $\ell(\cdot) \in \mathcal{L}_s$

$$E_F \ell(Y) = \int_0^\infty \{1 - F(y) + F(-y)\}\, \mathrm{d}\ell(y), \tag{3.7}$$

it is clear that (3.3) and (3.4) are equivalent to

$$E_G \ell(Y) \geq E_F \ell(Y), \quad \text{all } \ell(\cdot) \in \mathcal{L}_s. \tag{3.8}$$

Of course, folklore says that the mid-range is optimal in estimating the midpoint of a uniform distribution, but (3.2), (3.6) and (3.8) make this claim precise.

4. FINITE SAMPLE EFFICIENCY OF SAMPLE MINIMUM

First we will consider estimation of the starting point θ of an exponential density, i. e. we will take

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x), \quad x \in \boldsymbol{R}. \tag{4.1}$$

Here $c(\eta) = e^{\lambda\eta}$, $\eta \in \boldsymbol{R}$, and for $\theta \geq 0$ inequality (2.5) of Proposition 2.1 yields

$$1 - G_n(y - \theta) \leq e^{\lambda\theta} \left(1 - G_n(y)\right). \tag{4.2}$$

Choosing $0 < u < v < 1$, $y = G_n^{-1}(v)$, $\theta = \lambda^{-1} \log((1-u)/(1-v))$, this reduces to

$$1 - G_n\left(G_n^{-1}(v) - \theta\right) \leq \frac{1-u}{1-v}\left(1 - G_n\left(G_n^{-1}(v)\right)\right) \leq 1 - u,$$

and hence to

$$G_n^{-1}(v) - G_n^{-1}(u) \geq -\frac{1}{\lambda}\log(1-v) - \left(-\frac{1}{\lambda}\log(1-u)\right), \quad 0 < u < v < 1. \tag{4.3}$$

Defining the spread ordering between distributions F and G as

$$G \geq_1 F \quad \text{iff} \quad \left|G^{-1}(v) - G^{-1}(u)\right| \geq \left|F^{-1}(v) - F^{-1}(u)\right|, \quad 0 < u, v < 1, \tag{4.4}$$

we see that (4.3) means that G_n is at least as spread out as the exponential distribution F_λ with parameter λ, i. e.

$$G_n \geq_1 F_\lambda. \tag{4.5}$$

Consequently, for every $a \in \boldsymbol{R}$, $T_n = X_{(1)} + a$ is optimal in the spread ordering. Consider the class $\mathcal{L}$ of arbitrary (possibly asymmetric) loss functions $\ell(\cdot)$; so $\ell : \boldsymbol{R} \to [0,\infty)$ is nonincreasing– nondecreasing and vanishes at 0. As in (2.1) of Theorem 2.1 of Klaassen [5] we conclude

$$\inf_{T_n} E_\theta \ell\left(n(T_n - \theta)\right) = \inf_b \int_0^\infty \ell(y+b)\lambda e^{-\lambda y}\, dy, \tag{4.6}$$

and we see that $X_{(1)} + a$ is a minimum risk estimator with a depending on λ, n and the loss function $\ell(\cdot)$. Via a more complicated proof (4.5) was obtained already in Corollary 2.1 of Klaassen [6].

Next we will consider a semiparametric problem, namely estimation of the starting point of an arbitrary nonincreasing density. More specifically we take $f(\cdot)$ vanishing on $(-\infty, 0]$ and nonincreasing on $(0, \infty)$. Without the monotonicity constraint on the density this problem has been studied extensively in an asymptotic setting; see Smith [7], p. 1175, for references. Now $c(\eta) \leq 1$ for $\eta \leq 0$ and for $\theta \geq 0$, inequality (2.6) of Proposition 2.1 yields

$$\begin{aligned} &G_n(y+\theta) - G_n(y) \leq P_f(nX_{(1)} \leq \theta), \quad y \in \boldsymbol{R}, \quad \text{or} \\ &\sup_y P_f(y < nT_n \leq y + z) \leq P_f(0 < nX_{(1)} \leq z), \quad z \geq 0. \end{aligned} \tag{4.7}$$

Note that in particular

$$\sup_y P_f(y < nX_{(1)} \leq y + z) = P_f(0 < nX_{(1)} \leq z), \quad z \geq 0. \tag{4.8}$$

Let us introduce an ordering $\geq_m$ for distribution functions F via their maximum concentration function $\sup_y\{F(y+z) - F(y)\}$, $z \geq 0$, as we will call it. Then, (4.7) and (4.8) mean that the maximum concentration function of G_n equals at most that of F_n with $F_n(y) = P_f(nX_{(1)} \leq y)$, and we will denote this by

$$G_n \geq_m F_n. \tag{4.9}$$

Note that for every $a \in \boldsymbol{R}$, $T_n = X_{(1)} + a$ is an optimal estimator in this ordering, i. e. has a maximal maximum concentration function. Consequently, $[X_{(1)} - F^{-1}(1-(1-\alpha)^{1/n}), X_{(1)})$ is the optimal (shortest) confidence interval for θ with confidence coefficient $1 - \alpha$; here $F^{-1}(\cdot)$ denotes the quantile function for the density $f(\cdot)$. Consider the class $\mathcal{L}_\ell$ of loss functions which are constant at the left-hand side; $\ell : \boldsymbol{R} \to [0, \infty]$, $\ell(0) = 0$, there exists a $b_0 \geq 0$, such that $\ell(\cdot)$ is nondecreasing on $[-b_0, \infty)$ and constant on $(-\infty, -b_0)$ with $\lim_{x\to-\infty} \ell(x) = \lim_{x\to\infty} \ell(x)$. Then we may write

$$\ell(x) = \int_0^\infty \{1 - \mathbf{1}_{[-b_0,z)}(x)\}\, \mathrm{d}\ell(z), \quad x \in \boldsymbol{R},$$

and we see that (4.7) yields

$$\begin{aligned} \inf_b E_f \ell(nT_n - b) &= \inf_b \int_0^\infty \{1 - P_f(-b_0 + b \leq nT_n < z + b)\}\, \mathrm{d}\ell(z) \\ &\geq \int_0^\infty \left\{1 - \sup_y P_f\left(y \leq nT_n < y + (b_0 + z)\right)\right\} \mathrm{d}\ell(z) \\ &\geq \int_0^\infty \left\{1 - P_f\left(0 \leq nX_{(1)} < b_0 + z\right)\right\} \mathrm{d}\ell(z) \\ &= E_f \ell\left(nX_{(1)} - b_0\right), \end{aligned} \tag{4.10}$$

and in particular

$$\inf_b E_f \ell(nX_{(1)} - b) = E_f \ell(nX_{(1)} - b_0). \tag{4.11}$$

Finally, we note that $1 - \mathbf{1}_{[0,z)}(\cdot) \in \mathcal{L}_\ell$ and we conclude that (4.9) is equivalent to

$$\inf_b E_f \ell(nT_n - b) \geq \inf_b E_f \ell(nX_{(1)} - b), \quad \ell(\cdot) \in \mathcal{L}_\ell. \tag{4.12}$$

(Received August 30, 1994.)

REFERENCES

[1] N. H. Bingham: The sample mid-range and symmetrized extremal laws. Preprint. University of London 1994.
[2] J. Hájek: A characterization of limiting distributions of regular estimates. Z. Wahrsch. verw. Geb. *14* (1970), 323–330.
[3] J. Hájek: Local asymptotic minimax and admissibility in estimation. In: Proc. Sixth Berkeley Symp. Math. Statist. Probab. 1970, Vol I (L. M. Le Cam, J. Neyman and E. L. Scott, eds.), Univ. of California Press 1972, pp. 175–194.
[4] C. A. J. Klaassen: Statistical Performance of Location Estimators. (Mathematical Centre Tracts 133.) Mathematical Centre, Amsterdam 1981.
[5] C. A. J. Klaassen: Location estimators and spread. Ann. Statist. *12* (1984), 311–321.
[6] C. A. J. Klaassen: Estimators and spread. Ann. Statist. *17* (1989), 859–867.
[7] R. L. Smith: Estimating tails of probability distributions. Ann. Statist. *15* (1987), 1174–1207.
[8] C. J. Stone: Asymptotic properties of estimators of a location parameter. Ann. Statist. *2* (1974), 1127–1137.
[9] A. W. van der Vaart: An asymptotic representation theorem. Internat. Statist. Rev. *59* (1991), 97–121.

Prof. Dr. Chris A. J. Klaassen, Department of Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam. The Netherlands.