Comm. Korean Math. Soc. **13** (1998), No. 3, pp. 629–634

ON ROBUST MINIMAX APPROACH UNDER FINITE DISTRIBUTIONS

GEORGIY L. SHEVLYAKOV, JAE WON LEE AND SUNG WOOK PARK

ABSTRACT. As most of distributions appearing in applications are finite but with the unknown domain of finiteness, we propose to use the robust minimax approach for the determination of the boundaries of this domain. The obtained least favorable distribution minimizing Fisher information over the class of the approximately Gaussian finite distributions gives the reasonable sizes of the domain of finiteness and the thresholds of truncation.

1. Introduction

Robust methods are used to provide the stability of statistical inference under the departures from the accepted distribution model. One of the basic approaches to the synthesis of robust estimation procedures is the minimax principle. In this case, in a given class of distribution densities the least favorable one minimizing Fisher information is determined. The unknown parameters of a distribution model are then estimated by means of the maximum likelihood method for this density (Huber, [1]).

In our paper, we use the robust minimax approach to construct the definite boundaries of the considered class of the truncated distributions. Dealing with the real-life problems of data processing, a statistician usually has at his disposal some information on the natural boundaries of the data dispersion: the arbitrary large data values do not ever appear. As a rule, this information is of a very uncertain character. Our main goal is to present a possible way of the formalization of such information about distributions in order to design the precise rules for truncating the

---

Received July 7, 1997. Revised March 26, 1998.

1991 Mathematics Subject Classification: 62F35, 62F12.

Key words and phrases: robust minimax approach, location, finite distributions, truncated distributions.

This work was supported by the Kumoh National University of Technology.

distributions apriori defined in the infinite domain.

In Section 2, we present a brief survey of the main results within the robust minimax approach necessary for understanding our solution. In Section 3, we set the problem and obtain the main result. In Section 4, the proof is given.

2. Robust minimax estimation of a location parameter

Let $x_1, \cdots, x_n$ be data with $X_i \sim f(x-\theta)$. The M-estimator $\hat{\theta}$ of a location parameter θ is defined by Huber [2] as a zero of $\sum_{i=1}^{n} \psi(x_i - \cdot)$, where $\psi(\cdot)$ is a score function belonging to a certain class Ψ. The minimax approach implies the determination of the least favourable density f^* minimizing Fisher information $I(f)$ in a convex class $\mathcal{F}$

$$(1) \qquad f^* = \arg\min_{f \in \mathcal{F}} I(f), \qquad I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,dx,$$

followed by designing the optimal maximum likelihood estimator with a score function of the form

$$(2) \qquad \psi^*(x) = -f^{*\prime}(x)/f^*(x)\,.$$

Under rather general conditions put upon the classes $\mathcal{F}$ and Ψ (Huber [2, 3]), $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normally distributed and the asymptotic variance $V(\psi, f)$ has the saddle point (ψ^*, f^*) with the corresponding minimax property

$$V(\psi^*, f) \leq V(\psi^*, f^*) \leq V(\psi, f^*).$$

In robustness studies, the following conditions for the classes of distributions $\mathcal{F}$ are common:

$$(3) \qquad f(x) \geq 0, \quad f(-x) = f(x), \quad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Depending on the additional restrictions upon the class $\mathcal{F}$, different forms of the least favorable density f^* and the corresponding score function ψ^* may appear.

There are many results on the least favorable distributions in the different classes of ε-contaminated neighborhoods of a given distribution (see Huber [2, 3], Saks and Ylvisaker [4]). The class of the *approximately finite* distributions with the restriction upon the central part of a distribution (Huber [3]) is important for further considerations:

$$\mathcal{F} = \left\{ f \, : \, \int_{-l}^{l} f(x)\, dx \geq 1 - \beta, \quad 0 < \beta < 1 \right\},$$

where l and β are given parameters, the later characterizing the level of a prior uncertainty. The least favorable density consists of the *cosine*-type and the *exponential*-type parts (Huber, [3]):

$$f^*(x) = \begin{cases} A_1 \cos^2(B_1 x), & |x| \leq l, \\ A_2 \exp(-B_2 |x|), & |x| > l. \end{cases}$$

The constants A_1, A_2, B_1 and B_2 are determined from the system of equations including the norming condition, the characterizing restriction of the approximate finiteness, and the transversality conditions inducing the smooth gluing at $|x| = l$

$$\int_{-\infty}^{\infty} f^*(x)\, dx = 1, \quad \int_{-l}^{l} f^*(x)\, dx = 1 - \beta,$$
$$f^*(l-0) = f^*(l+0), \quad f^{*\prime}(l-0) = f^{*\prime}(l+0).$$

The remarkable feature of this robust solution (and also others, see Huber [3]) is the presence of the *exponential* "tails": it is due to the fact that the extremals of the basic variational problem 91) are exponents.

3. Problem statement and main result

We consider the new class of the approximately Gaussian finite distributions

$$f(x) = \begin{cases} A\phi(x), & |x| \leq l, \\ h(x), & l < |x| \leq L, \\ 0, & |x| > L, \end{cases} \tag{4}$$

EXAMPLE. With $\beta = 0.1$ we have the following numerical results:

$$\omega = 0.803\pi, \quad l = 2.809, \quad L = 3.497,$$

representing the reasonable thresholds of the truncation at the level of "$3.5\,\sigma$".

Finally, we note that the robust minimax estimator of a location parameter is given by (2) with the score function $\psi^*(x) = -f^{*\prime}(x)/f^*(x)$, and it provides the resistance to outliers in data rejecting them at the threshold $|x| = L$.

4. PROOF. First, we elucidate the structure of the solution (6) and then prove its optimality. The variational problem (1) with the side condition of norming is reformulated by the use of the following change of variables $f(x) = g^2(x) \geq 0$,

$$\text{(7)} \quad \text{minimize} \quad J(g) = \int_{-\infty}^{\infty} g'(x)^2\, dx \quad \text{subject to} \quad \int_{-\infty}^{\infty} g^2(x)\, dx = 1.$$

The Lagrange functional for the problem (7) is given by

$$L(g, \lambda) = \int_{-\infty}^{\infty} g'(x)^2\, dx + \lambda \left(\int_{-\infty}^{\infty} g^2(x)\, dx - 1 \right).$$

Then the Euler equation for it has the form

$$g''(x) - \lambda g(x) = 0,$$

and, respectively, its solutions of the *cosine*-type are the extremals of the problem (7) under finite distributions. The optimum solution of the original problem (1) with the restrictions (4) and (5) is the smooth "gluing" of the free *cosine*-type extremals and the Gaussian shape density $A\phi(x)$. The parameters of "gluing" A_1, A_2, B, l, and L are determined from the conditions of norming, the approximately finiteness, continuity and differentiability of the solution at $|x| = l$, and from the natural boundary conditions $f(L) = f'(L) = 0$ at the free boundaries $|x| = L$ [1].

We now check the optimality of the obtained solution. It is known that the density f^* belonging to a convex class $\mathcal{F}$ minimizes Fisher information if and only if

$$\left[\frac{d}{dt}I(f_t)\right]_{t=0} \geq 0\,, \tag{8}$$

where $f_t = (1-t)f^*+tf$, and f is an arbitrary density with $0 < I(f) < \infty$ [3]. The inequality (8) can be rewritten as

$$\int_{-\infty}^{\infty} (2\psi^{*\prime} - \psi^{*2})(f - f^*)\, dx \geq 0, \tag{9}$$

where $\psi^*(x)$ is the optimal score function (2). The direct evaluation of the left part of (9) confirms its validity. □

References

[1] I. M. Gelfand and S. V. Fomin, *Calculus of Variations,* tran. Richard. A. Silverman, Prentice-Hall, Inc, Englewood Cliffs, N. J., 1963.
[2] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist. **35** (1964), 73–101.
[3] ———, *Robust Statistics*, John Wiley, New York, 1981.
[4] J. Sacks and D. Ylvisaker, *A note on Huber's robust estimation of a location parameter*, Ann. Math. Statist. **43** (1972), 1068–1075.

Georgiy L. Shevlyakov
Department of Mathematics
St. Petersburg State Technical University
St. Petersburg, 195251, Russia

Jae Won Lee and Sung Wook Park
Department of Applied Mathematics
Kumoh National University of Technology
Kumi 730-701, Korea