J. Korea Soc. Math. Educ. Ser. B: Pure Appl. Math.
Volume 8, Number 2 (November 2001), Pages 79–85
ISSN 1226-0657

ON THE MINIMAX ROBUST APPROACH TO THE TRUNCATION OF DISTRIBUTIONS

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

ABSTRACT. As most of distributions in applications have a finite support, we introduce the class of finite distributions with the known shape of their central part and the unknown tails. Furthermore,we use the Huber minimax approach to determine the unknown characteristics of this class. We obtain the least informative distributions minimizing Fisher information for location in the classes of the truncated Gaussian and uniform distributions, and these results give the reasonable values of the thresholds of truncation. The properties of the obtained solutions are discussed.

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 densities the least informative (favorable) one minimizing Fisher information is determined. The unknown parameters of a distribution model are then estimated by the means of the maximum likelihood method for this density (see Huber [3], [4]). The robust minimax procedures provide a guaranteed level of the estimator's accuracy (measured by the supremum of an asymptotic variance) for any density in a given class.

Dealing with the real-life problems of data processing, a statistician usually has an information on the natural boundaries of data dispersion: the arbitrarily large data values do not ever appear. As a rule, this information about the distribution tails and their boundaries is rather uncertain.

Received by the editors March 16, 2001 and in revised form June 21, 2001.
2000 *Mathematics Subject Classification.* 62F35, 62E10.
Key words and phrases. minimax approach, robustness, finite distributions, truncated distributions.
This work was supported by the Kumoh National University of Technology, 1997.

Our main goal is to suggest a method formalizing such an information about distributions and applying the Huber minimax approach to design the precise rules for truncating the distributions possibly 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 including the important applications for the truncated Gaussian and uniform distributions, and give a proof of the theorem. In Section 4, the properties of the obtained solution are discussed.

2. MINIMAX ROBUST ESTIMATION OF A LOCATION PARAMETER

Let $x_1, \cdots, x_n$ be independent random variables with common density $f(x-\theta)$ in a convex class $\mathcal{F}$. Then the M-estimator $\hat{\theta}$ of a location parameter θ is defined by Huber [3] as a zero of $\sum_1^n \psi(x_i - \cdot)$ with a suitable score function ψ. The minimax approach implies the determination of the least informative distribution density f_0 minimizing Fisher information $I(f)$ in the class $\mathcal{F}$ where

$$f_0 = \arg\min_{f\in\mathcal{F}} I(f), \quad I(f) = \int_{-\infty}^{\infty} (f'(x)/f(x))^2 f(x)\,dx, \tag{1}$$

followed by designing the optimum maximum likelihood estimator with the score function

$$\psi_0(x) = -f_0'(x)/f_0(x). \tag{2}$$

Under rather general conditions (for details, see Huber [3], [4]), $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normally distributed and the asymptotic variance $V(\psi, f)$ has the saddle point (ψ_0, f_0) with the corresponding minimax property

$$V(\psi_0, f) \le V(\psi_0, f_0) \le V(\psi, f_0).$$

The following conditions are assumed for the classes of distributions $\mathcal{F}$:

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

Depending on the additional restrictions upon a class $\mathcal{F}$, different forms of the least informative density f_0 and the corresponding score function ψ_0 may appear.

There are many results on the least informative distributions in the different classes of ε-contaminated neighbourhoods of a given distribution (see Huber [3], [4]; Sacks and Ylvisaker [5]; Collins and Wiens [1]; Wiens [7]). The qualitatively other

types of distribution classes with a bounded variance were considered by Vil'chevskiy and Shevlyakov [6]. In the class

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

of the *approximately finite* distributions (in other words, the distributions with a bounded subrange, see Huber [4]), where l and β are given parameters, the latter characterizing the level of a prior uncertainty of a distribution, the least informative density consists of the *cosine*-type and the *exponential*-type parts:

$$f_0(x) = \begin{cases} A_1 \cos^2(B_1 x), & |x| \leq l, \\ A_2 \exp(-B_2|x|), & |x| > l. \end{cases}$$

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 glueing at $|x| = l$ such that

$$\int_{-\infty}^{\infty} f_0(x)\,dx = 1, \qquad \int_{-l}^{l} f_0(x)\,dx = 1 - \beta,$$

$$f_0(l-0) = f_0(l+0), \qquad f_0'(l-0) = f_0'(l+0).$$

The remarkable feature of this robust solution (and also others, see Huber [4]) is the presence of the *exponential* "tails": it is due to the fact that the extremals of the basic variational problem are exponents. In the case of $\beta = 0$, approximately finite distributions become finite

$$\mathcal{F} = \left\{ f : \int_{-l}^{l} f\,dx = 1 \right\},$$

and the least informative density is of the form

$$f_0(x) = \begin{cases} \cos^2(\pi x/(2l))/l, & |x| \leq l, \\ 0, & |x| > l. \end{cases}$$

3. Problem Statement and Main Result

A statistician usually has the more or less definite information about the central part of a distribution and the rather vague considerations about the tails. Moreover, most of distributions in applications seem to be finite but with the unknown domain

of finiteness. According to these qualitative considerations, we now introduce the following class $\mathcal{F}_{tr}$ of the truncated distributions

$$\mathcal{F}_{\mathrm{tr}} = \left\{ f:\ f(x) = \begin{cases} p(x), & |x| \le l; \\ h(x), & l < |x| \le L; \\ 0, & |x| > L \end{cases} \right\} \tag{4}$$

with the side conditions and with the restriction on the central part of a distribution

$$\int_{-l}^{l} f(x)\,dx = 1 - \beta, \quad 0 < \beta < 1, \tag{5}$$

where $p(x)$ is a given probability density in the central part; $h(x)$ is an arbitrary nonnegative symmetric function (the unknown tails); l and L are arbitrary constants: the latter L defines the domain of finiteness; the value of the parameter β is given, and, in this case, it is a measure of the uncertainty of our knowledge about the central part of a distribution: $0 < \beta < 1$, naturally small with $\beta = 0.05$ or $\beta = 0.1$. We also assume that $f(x)$ are continuously differentiable functions.

We suggest to determine all unknown characteristics of this distribution applying Huber minimax approach and minimizing Fisher information.

Theorm. *In the class $\mathcal{F}_{\mathrm{tr}}$, the least informative density is of the form*

$$f_0(x) = \begin{cases} p(x), & |x| \le l; \\ A\cos^2(B(|x| - x_0)/2), & l < |x| \le L; \\ 0, & |x| > L, \end{cases} \tag{6}$$

where the constants A, B, l and L are determined from the following system of equations

$$\int_{-L}^{L} f_0(x)\,dx = 1, \qquad \int_{-l}^{l} f_0(x)\,dx = 1 - \beta.$$

Remark 1. We have all the restrictions of the class of *approximately finite* distributions but the later additional boundary condition (the natural boundary condition) providing the finiteness of Fisher information $(0 < I(f) < \infty)$.

Example 1. With the Gaussian $p(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ we have the following relations:

$$A = \phi(l)(1 + \tan^2(\omega/2)), \quad B = l/\tan(\omega/2), \quad x_0 = l - (1/l)\omega\tan(\omega/2),$$
$$l = \Phi^{-1}(1 - \beta/2), \quad L = x_0 + (\pi/l)\tan(\omega/2),$$

where the value of the auxiliary parameter $\omega = B(l - x_0)$ is determined from

$$\omega = 2 \arctan \frac{[\beta + 2\phi(l)l]l}{\pi\phi(l)},$$

and $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$. Minimum information is

$$I(f_0) = \int_{-l}^{l} \phi(x)\, dx + AB^2(L - l) + AB \sin\omega.$$

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

$\omega = 2.297, A = 0.6132, B = 0.7391, x_0 = -1.463, l = 1.645, L = 2.798, I(f_0) = 1.285,$

representing the reasonable thresholds of truncating at the level of "3σ".

Example 2. With the uniform distribution

$$p(x) = \begin{cases} 0.5, & |x| \le 1; \\ 0, & |x| > 1 \end{cases}$$

we have:

$$A = 0.5,\ B = \pi/(2\beta),\ x_0 = l,\ l = 1 - \beta,\ L = 1 + \beta,\ I(f_0) = \pi^2/(4\beta).$$

Remark 2. In this case, the *cosine-type tails* regularize the discontinuous uniform distribution providing the finiteness of Fisher information.

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

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

The Lagrange functional for this problem 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 in the class $\mathcal{F}_{\text{tr}}$ (4) of the truncated distributions. The optimum solution of the original problem in the class $\mathcal{F}_{\text{tr}}$ is the smooth "gluing" of the free cosine-type extremals and the given density $p(x)$. The parameters of "gluing" A, B, x_0, l, and L are determined from the conditions (3) and (5), continuity and differentiability of the solution at $|x| = l$, and

from the natural boundary condition $f'(L) = 0$ at the free boundaries $|x| = L$ (see Gelfand and Fomin [2]).

We now check the optimality of the obtained solution. It is known (see Huber [4]) that the density f_0 belonging to a convex class $\mathcal{F}$ minimizes Fisher information if and only if

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

where $f_t = (1-t)f_0 + tf$, and f is an arbitrary distribution density providing $0 < I(f) < \infty$. This inequality can be rewritten as

$$\int_{-\infty}^{\infty} (2\psi_0' - \psi_0^2)(f - f_0)\,dx \geq 0,$$

where $\psi_0(x)$ is the optimal score function (2). The direct evaluation of the lefthand side of (7) concludes the proof. □

4. FINAL REMARKS

First, we present a few additional considerations on the choice of the class $\mathcal{F}_{\text{tr}}$. The least informative density f_0 in the class of ε-contaminated distributions (see Huber [3]) has the known central part: $f_0(x) = (1-\varepsilon)p(x)$ and exponential tails, but with other ε-neighbourhoods of a known density $p(x)$ the central part of f_0 may be of a rather exotic shape differing much from $p(x)$ (see Huber [4], Wiens [7]). So, it seems more realistic to postulate the knowledge of a central zone of a distribution and to attribute all uncertainty to distribution tails. We repeat once more that the assumption of finiteness gives a more adequate distribution model within many applications. In the class $\mathcal{F}_{\text{tr}}$, it is no need in the assumption of strong unimodality of $p(x)$ (see Huber [3]) and its weakened variants (see Wiens [7]): $p(x)$ can be of an arbitrary form, for example, U-shaped. Certainly, symmetry is necessary. The natural structure of the class $\mathcal{F}_{\text{tr}}$ inherently corresponds to the expected form of f_0, thus it gives the possibility to obtain a rather simple solution.

Second, we comment on the properties of the minimax M-estimator of location. The optimal score function ψ_0 provides the guaranteed level of accuracy of the minimax estimator: $V(\psi_0, f) \leq V(\psi_0, f_0)$ for all $f \in \mathcal{F}_{\text{tr}}$ and its robustness by rejecting outliers in data at the threshold $|x| = L$. We indicate that the values of Fisher information with truncated distributions are larger than with the original ones (see Example 1: $I(\phi) = 1 < 1.285 = I(\phi_{\text{tr}})$). Moreover, the minimax estimators

are *superefficient*: the integrals diverge in the expression for the asymptotic variance $V(\psi_0, f) = \int \psi_0^2 f dx / [\int \psi_0' f dx]^2$ in such a way that $V(\psi_0, f) = 0$ for $f \notin \mathcal{F}_{\mathrm{tr}}$. Thus, using ψ_0 with large samples, we have highly precise estimates of location.

Finally, we note that the small size sample behaviour of the minimax estimators is unknown, and the rule of rejection of outliers can be based on the thresholds l and L.

REFERENCES

1. Collins, J. and Wiens, D.: Minimax variance M-estimators in ε-contamination models. *Ann. Statist.* **13** (1985), no. 3, 1078–1096. MR **87h:**62062
2. Gelfand, I. and Fomin, S.: *Calculus of Variations.* Revised English edition, translated and edited by Richard A. Silverman. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963. MR **28#**3353
3. Huber, P.: Robust estimation of a location parameter. *Ann. Math. Statist.* **35** (1964) 73–101. MR **28#**4622
4. _____: *Robust Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1981. MR **82i:**62057
5. Sacks, J. and Ylvisaker, D.: A note on Huber's robust estimation of a location parameter. *Ann. Math. Statist.* **43** (1972), 1068–1075. MR **47#**6023
6. Vil'chevskiy, N. and Shevlyakov, G.: Robust minimax estimation of a location parameter with a bounded variance. In: *Proceedings of the XV Seminar on Stability Problems for Stochastic Models*, June 2–6, 1992, Perm, Russia, edited by V. M. Zolotarev et al. (pp. 279–288). TVP/VSP, Moscow/Utrecht, 1994.
7. Wiens, D.: Minimax variance M-estimators of location in Kolmogorov neibbourhoods. *Ann. Statist.* **14** (1986), 724–732. MR **87m:**62102

(J. W. LEE) DEPARTMENT OF APPLIED MATHEMATICS, KUMOH NATIONAL UNIVERSITY OF TECHNOLOGY, 188 SINPYEONG-DONG, GUMI, GYEONGBUK 730-701, KOREA
E-mail address: ljaewon@knut.kumoh.ac.kr

(G. L. SHEVLYAKOV) DEPARTMENT OF MATHEMATICS, ST. PETERSBURG STATE TECHNICAL UNIVERSITY, ST. PETERSBURG, RUSSIA
E-mail address: shev@stat.hop.stu.neva.ru

(S. W. PARK) DEPARTMENT OF APPLIED MATHEMATICS, KUMOH NATIONAL UNIVERSITY OF TECHNOLOGY, 188 SINPYEONG-DONG, GUMI, GYEONGBUK 730-701, KOREA
E-mail address: swpark@knut.kumoh.ac.kr