J. Korean Math. Soc. **35** (1998), No. 3, pp. 755–764

ON THE LEAST INFORMATIVE DISTRIBUTIONS UNDER THE RESTRICTIONS OF SMOOTHNESS

JAE WON LEE, SUNG WOOK PARK, NIKITA VIL'CHEVSKIY AND GEORGIY SHEVLYAKOV

ABSTRACT. The least informative distributions minimizing Fisher information for location are obtained in the classes of continuously differentiable and piece-wise continuously differentiable densities with the additional restrictions on their values at the median and mode of population in the point and interval forms. The structure of these optimal solutions depends both on the assumptions of smoothness and form of characterizing restrictions of the class of distributions: in the class of continuously differentiable densities, the least informative distributions are finite and have the *cosine*-type form, and, in the class of piece-wise continuously differentiable densities, the least informative densities have *exponential*-type tails, the Laplace density in particular. The dependence of optimal solutions on the assumptions of symmetry is also analyzed.

1. Introduction

The least informative (favorable) distributions minimizing Fisher information appear within the minimax robust approach proposed by Huber [1]. 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 one minimizing Fisher information is determined. The unknown parameters of a distribution model are then estimated by

Received February 24, 1998.
1991 Mathematics Subject Classification: Primary 62F35, secondary 62B10.
Key words and phrases: minimax approach, robustness, least informative distributions, location parameter.
This work was supported by the Kumoh National University of Technology Research Funds.

the means of the maximum likelihood method for this density. 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.

Let $x_1, \dots, x_n$ 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 as a solution of the following equation

$$\sum_{i=1}^{n} \psi(x_i - \hat{\theta}) = 0$$

with a suitable score function ψ.

The minimax approach implies the determination of the least informative density f_0 minimizing Fisher information $I(f)$ in the class $\mathcal{F}$

$$f_0 = \arg\min_{f \in \mathcal{F}} I(f), \quad I(f) = \int \left[\frac{f'(x)}{f(x)}\right]^2 f(x)\, dx, \tag{1}$$

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

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

Under rather general conditions (see Huber, [1]), $\sqrt{n}(\hat{\theta}-\theta)$ 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 left-hand part of this inequality shows that the minimax variance of the M-estimator with the score function ψ_0 provides the guaranteed level of the estimator's accuracy for any density in the class $\mathcal{F}$:

$$V(\psi_0, f) \le V(\psi_0, f_0) = \frac{1}{n\, I(f_0)}.$$

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

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

Depending on the additional, characterizing the class $\mathcal{F}$, restrictions, different forms of the least informative density f_0 and the corresponding score function ψ_0 may result. Note that the solution of the variational problem (1) with the only side restrictions (3) is degenerate: $f_0(x) \equiv 0$ for all $x \in \mathbf{R}$.

It is a well-known fact that the properties of the solutions of variational problems depend on the assumptions of smoothness, and our main goal is to show that the character of these assumptions determine the essential features of least informative distributions, for instance, they determine whether these distributions are finite or not.

There are many results on the least informative distributions under various characterizing restrictions, most of them are concerned with the classes of ε-contaminated neighborhoods of a given distribution (see Huber, [1]; Sacks and Ylvisaker, [2]; Tsypkin, [3]; Collins and Wiens, [4]; Wiens, [5]). The qualitatively other types of distribution classes with a bounded variance were considered by Vil'chevskiy and Shevlyakov [6]. The common and basic feature of these solutions is the presence of *exponential* tails, and they imply the boundness of the corresponding score functions and robustness of M-estimators of location.

We now are especially interested in the following solutions. In the class of the *approximately finite* distributions with the bounded subrange (see Huber, [1])

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

where l and β are given parameters of this class, 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:

$$f_0(x) = \begin{cases} A_1 \cos^2(B_1 x), & |x| \leq l, \\ A_2 \exp(-B_2 |x|), & |x| > l. \end{cases} \tag{4}$$

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 (see Gelfand and Fomin, [7]) inducing the smooth glueing at $|x| = l$

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

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

In the case of $\beta = 1$, approximately finite distributions become finite

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

and the least informative density is of the form

$$f_0(x) = \begin{cases} l^{-1} \cos^2(\pi x/(2l)), & |x| \le l, \\ 0, & |x| > l. \end{cases} \tag{5}$$

In the class of nondegenerate densities (see Tsypkin, [3])

$$\mathcal{F} = \left\{ f: \quad f(0) \ge \frac{1}{2d} > 0 \right\},$$

the least informative density is the Laplace

$$f_0(x) = \frac{1}{2d} \exp\left(-\frac{|x|}{d} \right). \tag{6}$$

The same structures also manifest in the more general cases.

2. Least informative distributions: the case of continuously differentiable densities

The information on the basic characteristics of distributions such as mode, median, expectation, variance, etc., can be obtained by applying the methods of exploratory data analysis. It seems naturally that this information should be taken into account while designing methods and algorithms of data processing, for instance, within the minimax robust approach.

THEOREM 1. *Let the value of a density $f(x)$ at mode $Mo = \arg\max f(x)$ be given: $f(Mo) = f_m$. Suppose $f(x)$ is continuously differentiable on* **R**. *Then the least informative density f_0 minimizing Fisher information for location is given by:*

$$f_0(x) = \begin{cases} f_m \cos^2[\pi f_m (x - Mo)/2], & |x - Mo| \le 1/f_m, \\ 0, & |x - Mo| > 1/f_m. \end{cases} \tag{7}$$

COROLLARY 1. *The least informative density* (7) *is also valid under the restriction of the inequality form*

$$f(Mo) \ge f_m > 0. \tag{8}$$

The similar results are valid under the restrictions on the value of a density at median of population.

THEOREM 2. *Let the value of a density $f(x)$ at median $Me = F^{-1}(1/2)$ be given: $f(Me) = f_m$. Suppose $f(x)$ is continuously differentiable on* $\mathbf{R}$. *Then the least informative density f_0 minimizing Fisher information for location is given by:*

$$f_0(x) = \begin{cases} f_m \cos^2[\pi f_m (x - Me)/2], & |x - Me| \le 1/f_m, \\ 0, & |x - Me| > 1/f_m. \end{cases} \tag{9}$$

COROLLARY 2. *The least informative density* (9) *is also valid under the restriction of the inequality form*

$$f(Me) \ge f_m > 0. \tag{10}$$

Minimum information is

$$I(f_0) = \pi^2 f_m^2.$$

3. Least informative distributions: the case of piece-wise continuously differentiable densities

We now consider the case of piece-wise continuously differentiable densities under the same conditions as in Section 2.

THEOREM 3. *Let the values of a density at mode or at median be given:*

$$f(Mo) = f_m \quad \text{or} \quad f(Me) = f_m.$$

Suppose $f(x)$ is piece-wise continuously differentiable on $\mathbf{R}$. *Then the least informative density f_0 minimizing Fisher information is the Laplace:*

$$f_0(x) = f_m \exp\left(-2 f_m |x - x_0|\right), \tag{11}$$

where $x_0 = Mo$ or $x_0 = Me$, respectively.

COROLLARY 3. *The Laplace density* (11) *is also the least informative under the restrictions of the inequality form* (8) *and* (10).

Minimum information is

$$I(f_0) = 4 f_m^2.$$

The obtained results can be generalized on the cases of integral restrictions on the mass of a distribution in the central zone:

$$\int_a^b f(x)\,dx = 1-\beta > 0, \quad Mo \in (a,b) \tag{12}$$

and

$$\int_a^b f(x)\,dx = 1-\beta > 0, \quad Me \in (a,b), \tag{13}$$

where a, b and β $(0<\beta<1)$ are given.

THEOREM 4. *Under the conditions* (12), (13) *and with piece-wise continuously differentiable densities on* $\mathbf{R}$, *the least informative density* f_0 *minimizing Fisher information for location is given by formula* (4)*:*

$$f_0(x) = \begin{cases} A_1 \cos^2[B_1(x-x_0)], & |x-x_0| \le l, \\ A_2 \exp(-B_2|x-x_0|), & |x-x_0| > l, \end{cases} \tag{14}$$

where $x_0 = (a+b)/2$ *and* $l = (b-a)/2$. *The constants* A_1, A_2, B_1 *and* B_2 *are determined from the relations*

$$A_1 = \frac{(1-\beta)\delta}{(\delta+\sin\delta)l}, \qquad B_1 = \frac{\delta}{2l},$$

$$A_2 = \frac{\beta\lambda e^{\lambda}}{2l}, \qquad B_2 = \frac{\lambda}{2l},$$

where the auxiliary parameter δ *is the solution of the following equation*

$$\frac{(\delta+\sin\delta)\tan\delta}{2(\cos\delta)^2} = \frac{1-\beta}{\beta},$$

and λ *is given by*

$$\lambda = \delta\tan\delta.$$

COROLLARY 4. *The least informative density* (14) *is also valid under the integral restrictions of the inequality form*

$$\int_a^b f(x)\,dx \ge 1-\beta > 0, \quad Mo \in (a,b) \quad \textit{or} \quad Me \in (a,b).$$

Minimum information is

$$I(f_0) = \frac{4\lambda\delta^2}{(\lambda+1)l^2}.$$

REMARK 1. Note that all obtained least informative distributions are symmetric, though no any assumptions of symmetry have been assumed.

4. Least informative distributions: the dependence on symmetry assumptions

Now we consider the dependence of optimal solutions on the assumptions of symmetry.

THEOREM 5. *Let mode Mo be in $[a, b]$, median Me – in $[c, d]$, and $f(Mo) = f_m$. The parameters a, b, c, d and f_m of this class are assumed given. Let $f(x)$ be in the class of continuously differentiable functions. Under these assumptions, the least informative density f_0 minimizing Fisher information for location is given by:*

- *if* $[a, b] \bigcap [c, d] \neq \emptyset$ *then*

$$f_0(x) = \begin{cases} f_m \cos^2(\pi f_m (x - x_0)/2), & |x - x_0| \leq 1/f_m, \\ 0, & |x - x_0| > 1/f_m, \end{cases}$$

where x_0 is an arbitrary point in $[a, b] \bigcap [c, d]$, and $I(f_0) = \pi^2 f_m^2$;

- *if* $[a, b] \bigcap [c, d] = \emptyset$ *then $f_0(x)$ is nonsymmetric with the similar cosine-type extremals and has a rather cumbersome structure (here we only announce this result — it needs a separate consideration).*

REMARK 2. Note that these least informative distributions are finite.

But if we assume the piece-wise continuously differentiability for $f(x)$ then the least informative density is described by *exponential*-type extremals and now is *not finite*.

THEOREM 6.

- *If* $[a, b] \bigcap [c, d] \neq \emptyset$ *then the least informative distribution is the Laplace:*

$$f_0(x) = f_m \exp(-2 f_m |x - x_0|),$$

where x_0 is an arbitrary point in $[a, b] \bigcap [c, d]$, and $I(f_0) = 4 f_m^2$;

- *if* $[a,b]\bigcap[c,d]=\emptyset$ *and* $b<c$ *then*

$$f_0(x)=\begin{cases} f_m\exp[2f_m\,\omega(x-b)], & x\le b,\\ f_m\exp[-2f_m\,\omega(x-b)], & b<x\le c,\\ f_m\exp[-2f_m\,\omega(c-b)]\exp[-2f_m\,\mu(x-c)], & x>c;\end{cases}$$

the parameters ω *and* μ *are obtained from the following equations*

$$\omega=2-\exp[-2f_m\,\omega(c-b)],\qquad \mu=\exp[-2f_m\,\omega(c-b)],$$

and

$$I(f_0)=4\{1+[1-\exp(-2\omega(c-b))]^2\}\,f_m^2.$$

REMARK 3. It follows from Theorem 5 and Theorem 6 that, under weak departures from symmetry assumptions, the least informative distributions still remain symmetric.

5. Proofs

Using variational methods, we first obtain the structure of the optimal solutions and then verify their validity.

Proof of Theorem 1. In this case, the variational problem (1) is written as

$$F(f_m)=\min\int_{-\infty}^{\infty}\left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,dx \tag{15}$$

subject to

$$f'(Mo)=0,\quad f(Mo)=f_m>0,\quad f(x)\ge 0,\quad \int_{-\infty}^{\infty}f(x)\,dx=1.$$

Consider a new function

$$f(x)=f_m u^2(f_m(x-Mo)),\quad u(t)\ge 0.$$

Then the variational problem (15) can be reformulated as

$$F(f_m)=4f_m^2J,$$

where J is a solution of the following variational problem:

$$J=\min\int_{-\infty}^{\infty}(u'(t))^2\,dt$$

subject to

$$u'(0) = 0, \quad u(0) = 1, \quad u(t) \geq 0, \quad \int_{-\infty}^{\infty} u^2(t)\, dt = 1.$$

Lagrange functional for this problem is given by

$$L(u, \lambda) = \int u'(t)^2\, dt + \lambda \left(\int u^2(t)\, dt - 1 \right),$$

where λ is Lagrange multiplier. Hence Euler equation has the form

$$u''(x) + \lambda u = 0, \tag{16}$$

and its solution satisfying to initial conditions has the form

$$u(t) = \cos \sqrt{\lambda} t, \quad \text{if} \quad |t| \leq \frac{\pi}{2\sqrt{\lambda}},$$

smoothly "glued" with "zero" density.

We now check the optimality of the obtained solution. It is known (see Huber, [1]) that the density f_0 minimizes Fisher information in a convex class $\mathcal{F}$ if and only if

$$\left[\frac{d}{d\varepsilon} I(f_\varepsilon) \right]_{\varepsilon=0} \geq 0, \tag{17}$$

where $f_\varepsilon = (1-\varepsilon) f_0 + \varepsilon f$, and f is an arbitrary density with $I(f) < \infty$.

This inequality can be rewritten as

$$\int_{-\infty}^{\infty} (2{\psi_0}' - {\psi_0}^2)(f - f_0)\, dx \geq 0, \tag{18}$$

where $\psi_0(x)$ is the optimal score function (2).

Finally, the direct evaluation of the left-hand side of (18) gives

$$\int [f(x) - f_0(x)]\, dx,$$

which by norming condition is evidently zero, and this remark concludes the proof.

Proof of Corollary 1. Let $f(Mo) = y \geq f_m > 0$, then $I(f_0) = \pi^2 y^2 \geq \pi^2 f_m^2$, and evidently minimum is attained at $y = f_m$.

Proof of Theorem 2. In main features, this proof repeats the above-described scheme.

Proof of Theorem 3. Another solution of Euler equation (16) is given by exponent $u(t) = e^{-|\sqrt{-\lambda} t|}$, therefore we have the Laplace density (11) as the least informative. Final check should be done using the inequality (18).

All remained proofs can be directly performed by checking the sign of the left-hand side of the inequality (18).

References

[1] Huber, P., *Robust Statistics*, John Wiley, New York, 1981.
[2] Sacks, J. and D. Ylvisaker, *A note on Huber's robust estimation of a location parameter*, Ann. Math. Statist. **43** (1972), 1068-1075.
[3] Tsypkin, Ya. Z., *Optimization under uncertainty*, Doklady Akad. Nauk SSSR **228** (1976), 1306-1309 (in Russian).
[4] Collins, J. and D. Wiens, *Minimax variance M-estimators in ε-contamination models*, Ann. Statist. **13** (1985), 1078-1096.
[5] Wiens, D., *Minimax variance M-estimators of location in Kolmogorov neibbourhoods*, Ann. Statist. **14** (1986), 724-732.
[6] Vil'chevskiy, N. and G. Shevlyakov, *Robust minimax estimation of a location parameter with a bounded variance*, In: Stability Problems for Stochastic Models, eds. V. M. Zolotarev et al., Moscow/Utrecht, TVP/VSP, pp. 279-288, 1994.
[7] Gelfand, I. and S. Fomin. *Calculus of Variations*, tran. Richard A. Silverman, Prentice-Hall Inc., Englewood Cliffs, N.J., 1963.

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

Nikita Vil'chevskiy and Georgiy Shevlyakov
St. Petersburg, Russia