Bull. Korean Math. Soc. **39** (2002), No. 1, pp. 23–31

ON THE MINIMAX VARIANCE ESTIMATORS OF SCALE IN TIME TO FAILURE MODELS

JAE WON LEE AND GEORGY L. SHEVLYAKOV

ABSTRACT. A scale parameter is the principal parameter to be estimated, since it corresponds to one of the main reliability characteristics, namely the average time to failure. To provide robustness of scale estimators to gross errors in the data, we apply the Huber minimax approach in time to failure models of the statistical reliability theory. The minimax variance estimator of scale is obtained in the important particular case of the exponential distribution.

1. Introduction

This work is concerned with the application of robust minimax approach to the traditional problems of the statistical reliability theory. Robust methods were proposed in the pioneer works of J. Tukey (1960), P. Huber (1964), and F. Hampel (1968), have been intensively developed since the seventies and rather definitely formed by present in the field of mathematical statistics. The basic reason of investigations in this field is of a general mathematical character. They are induced by the necessity of study of the optimal decisions stability with respect to possible deviations from the optimality assumptions. "Optimality" and "stability" are the mutually complementary characteristics of any mathematical procedure. It is well-known that the efficiency rate of many optimal decisions is rather sensible to "small deviations" from initial assumptions.

In mathematical statistics, the remarkable example of a such unstable optimal procedure is given by the least squares method: the efficiency of its estimates sharply decreases under some models of deviations from the normal distribution ([3], [4], [5]). The similar situation occurs in estimation of a scale parameter of the exponential distribution. The

Received April 6, 2001.

2000 Mathematics Subject Classification: 2F35, 62E10.

Key words and phrases: minimax approach, robustness, scale parameter, reliability.

This work was supported by the Kumoh National University of Technology 1997.

efficient maximum likelihood estimator of the failure intensity parameter λ is the inverse value of the sample mean of the time to failure data observations: $\widehat{\lambda} = 1/\overline{T}$, $\overline{T} = n^{-1}\sum t_i$. The linear structure of the sample mean type estimators results in their instability to the occasional appearance of rare outliers in the data. From the statistical point of view, this instability causes a sharp loss of the estimator efficiency under small deviations from the accepted data distribution model which in turn may lead to gross errors in designed reliability characteristics.

At present, there exist a lot of robust versions of various statistical methods providing the stability of statistical inference in data analysis. However, main results in robust statistics refer to the normal and its neighborhood models of distributions which do not usually fit time to failure data.

2. M-estimators of a scale parameter

Consider a sample $t_1, ..., t_n$ of i.i.d. random variables from the gross error model

(1) $$\mathcal{F}_1 = \{f : f(t;T) = (1-\varepsilon)T^{-1}g(t/T) + \varepsilon h(t), \quad 0 \le \varepsilon < 1\},$$

where $g(\cdot)$ is a given density; ε is a contamination parameter characterizing the level of uncertainty of the accepted $g(t)$ density model; $h(t)$ is an arbitrary density; T is a scale parameter under estimation. Then M-estimators of a scale parameter T are defined by Huber ([5]) as

(2) $$\sum_{i=1}^{n} \chi\left(t_i/\widehat{T}_n\right) = 0\,,$$

where χ is a score function. In the particular case of the completely known density $T^{-1}f(t/T)$, the score function

(3) $$\chi(z) = -z\, f'(z)/f\,(z) - 1$$

gives the maximum likelihood estimator of scale as the solution of equation (2). Under general conditions of regularity put on the densities f and on the score functions χ ([3], pp. 125, 139–140), the estimators $\widehat{T}_n$ are consistent, asymptotically normal and asymptotically efficient, and possess the minimax property with respect to the asymptotic variance $D\widehat{T}_n = n^{-1}V(\chi, f)$:

(4) $$V(\chi^*, f) \le V(\chi^*, f^*) \le V(\chi, f^*)\,,$$

where f^* is a least informative (favorable) density minimizing the Fisher information $I(f)$ for scale in class (1)

$$f^* = \arg\min_{f\in\mathcal{F}} I(f), \qquad I(f) = \int_0^\infty z^2 \left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz - 1\,; \tag{5}$$

χ^* is determined from equation (3) ([4]). Condition (4) defines the saddle-point (χ^*, f^*) of the asymptotic variance $V(\chi, f)$. The left-hand part of inequality (4) is of a practical importance: the choice of estimate (2) with the score function χ^* provides the guaranteed level of the accuracy of an estimator for all densities belonging to class (1):

$$V(\chi^*, f) \le V(\chi^*, f^*) = 1/(n\, I(f^*)). \tag{6}$$

3. Minimax variance estimators of scale in the gross error model

The problem of the design of the minimax estimator $\widehat{T}_n$ comes formally to variational problem (5) with the side condition of norming and the restricted characterization of class (1)

$$f^* = \arg\min_{f\in\mathcal{F}} I(f), \tag{7}$$

$$\int_0^\infty f(z)\,dz = 1, \tag{8}$$

$$f(z) \ge (1-\varepsilon)\, g(z). \tag{9}$$

The restricted characterization of class (1) is written down in the inequality form (9), apparently, it includes the non-negativeness condition.

THEOREM 1. *For continuously differentiable on* $(0,\infty)$ *and logarithmic convex densities* f *satisfying conditions* (8) *and* (9), *the solution of problem* (7) *is of the form:*

$$f^*(z) = \begin{cases} C_1 z^{k_1} & \text{for} \quad 0 < z \le \Delta_1, \\ (1-\varepsilon)\, g(z) & \text{for} \quad \Delta_1 < z \le \Delta_2, \\ C_2 z^{k_2} & \text{for} \quad z > \Delta_2, \end{cases} \tag{10}$$

where the constants C_1, Δ_1, k_1, C_2, Δ_2, *and* k_2 *are determined from the following equations:*

$$C_1 = (1-\varepsilon) g(\Delta_1)(\Delta_1)^{-k_1}\,, \quad k_1 = \Delta_1\, g'(\Delta_1)/g(\Delta_1),$$

$$C_2 = (1-\varepsilon) g(\Delta_2)(\Delta_2)^{-k_2}\,, \quad k_2 = \Delta_2\, g'(\Delta_2)/g(\Delta_2), \tag{11}$$

$$k_1 + k_2 = -2,$$

$$\frac{\Delta_1\, g(\Delta_1)}{k_1+1} + \int_{\Delta_1}^{\Delta_2} g(z)\, dz - \frac{\Delta_2\, g(\Delta_2)}{k_2+1} = \frac{1}{1-\varepsilon}.$$

Proof. First we obtain the structure of solution (9), and then prove its optimality. By using the change of variables $f(z) = g^2(z) \geq 0$, we rewrite variational problem (7) with side condition (8) as

$$J(g) = \int_0^\infty z^2 g'(z)^2\, dz \to \min, \qquad \int_0^\infty g^2(z)\, dz = 1. \tag{12}$$

The Lagrange functional for this problem has the form

$$L(g,\lambda) = \int_0^\infty z^2 g'(z)^2\, dz + \lambda \left(\int_0^\infty g^2(z)\, dz - 1 \right).$$

Hence the Euler equation is

$$z^2 g''(z) + 2zg'(z) - \lambda g(z) = 0,$$

and its solutions are the extremals of problem (12)

$$f_1(z) = g_1^2(z) = C_1 z^{k_1}, \quad f_2(z) = g_2^2(z) = C_2 z^{k_2}, \quad k_1 + k_2 = -2. \tag{13}$$

The structure of the optimal solution of problem (7) with side conditions (8) and (9) is given by smooth "gluing" of free extremals (13) with the restriction curve $(1-\varepsilon)g(z)$ in the form (9). The parameters of "gluing" $C_1, \Delta_1, k_1, C_2, \Delta_2,$ and k_2 are determined from the conditions of continuity and differentiability of the optimal solution at the points $z = \Delta_1$ and $z = \Delta_2$

$$f^*(\Delta_1 - 0) = f^*(\Delta_1 + 0), \qquad f^{*\prime}(\Delta_1 - 0) = f^{*\prime}(\Delta_1 + 0),$$

$$f^*(\Delta_2 - 0) = f^*(\Delta_2 + 0), \qquad f^{*\prime}(\Delta_2 - 0) = f^{*\prime}(\Delta_2 + 0),$$

which along with relation (13) and the norming condition yield system (11). The assumption of logarithmic convexity is sufficient to provide the existence of at most two points of intersection of free extremals with the restriction curve.

It is known ([4], p. 82 and Subsection 5.6) that the density f^* belonging to the convex class $\mathcal{F}$ minimizes the Fisher information if and only if

$$\left[\frac{d}{ds} I(f_s) \right]_{s=0} \geq 0, \tag{14}$$

where $f_s = (1-s)f^* + sf$, and f is an arbitrary density with $I(f) < \infty$. Inequality (14) can be rewritten as

$$(15) \qquad \int_0^\infty (2z\chi^{*\prime}(z) - \chi^{*2}(z))(f(z) - f^*(z))\, dz \geq 0,$$

where $\chi^*(z)$ is the optimal score function [4]. By evaluating the left-hand part of (15), it can be shown that it is equivalent to the characterization restriction of the variational problem,

$$f(z) - (1-\varepsilon)\, g(z) \geq 0$$

and this remark concludes the proof. □

The robust minimax variance estimator $\widehat{T}_n$ evaluated from equation (2) is completely determined by the following score function;

$$(16) \qquad \chi^*(z) = -z\frac{(f^*(z))'}{f^*(z)} - 1 = \begin{cases} -k_1 - 1, & 0 < z \leq \Delta_1, \\ -z\, g'(z)/g(z) - 1, & \Delta_1 < z \leq \Delta_2, \\ -k_2 - 1, & z > \Delta_2. \end{cases}$$

Consider the following notations:

$I_1 = \{i : t_i/\widehat{T}_n \leq \Delta_1\}$, $I_2 = \{i : t_i/\widehat{T}_n > \Delta_2\}$, $I = \{i : \Delta_1 < t_i/\widehat{T}_n \leq \Delta_2\}$

Then equation (2) can be written as

$$(17) \quad \sum_{i\in I_1}(-k_1 - 1) + \sum_{i \in I}\left(-\frac{t_i}{\widehat{T}_n}\frac{g'(t_i/\widehat{T}_n)}{g(t_i/\widehat{T}_n)} - 1\right) + \sum_{i\in I_2}(-k_2 - 1) = 0.$$

Denoting the numbers of observations belonging to the sets I_1, I_2 and I as n_1, n_2 and $n - n_1 - n_2$, respectively, we get from (17) that the structure of the robust minimax estimator with score function (16) is the structure of the trimmed maximum likelihood estimator with the n_1-deleted smallest and n_2-greatest observations, the rest of the sample being processed by the maximum likelihood method. In the limit case as $\varepsilon \to 1$, this estimator is the sample median

$$(18) \qquad \widehat{T}_n = \text{med}\, t = \begin{cases} t_{(h+1)}, & n = 2h+1, \\ (t_{(h)} + t_{(h+1)})/2, & n = 2h. \end{cases}$$

With $\varepsilon = 0$, the robust minimax variance estimator $\widehat{T}_n$ is the maximum likelihood estimator for a scale parameter of the density $g(t)$.

REMARK 1. The obtained results can be used for designing robust minimax estimators of scale for the Gamma, Weibull and exponential distributions, the first two with given parameters of form. The important case of the exponential distribution is analyzed below.

4. Example: robust minimax variance estimator of scale for the exponential distribution

THEOREM 2. *For continuously differentiable on* $(0,\infty)$ *densities* $f(z)$ *satisfying conditions* (8) *and* (9) *with* $g(z) = \exp(-z)$, *the solution of problem* (7) *is of the form: with* $0 \le \varepsilon < \varepsilon_0 = (1+e^2)^{-1}$,

$$(19) \qquad f^*(z) = \begin{cases} (1-\varepsilon)\, e^{-z} & \text{for } \; 0 < z \le \Delta, \\ C\, z^k & \text{for } \; z > \Delta, \end{cases}$$

where the constants C, k *and* Δ *are the functions of the parameter* ε

$$(20) \qquad C = (1-\varepsilon)e^{-\Delta}\Delta^{\Delta}, \quad k = -\Delta, \quad \frac{e^{-\Delta}}{\Delta - 1} = \frac{\varepsilon}{1-\varepsilon};$$

with $\varepsilon_0 \le \varepsilon < 1$

$$(21) \qquad f^*(z) = \begin{cases} C_1 z^{k_1} & \text{for } \; 0 < z \le \Delta_1, \\ (1-\varepsilon)\, e^{-z} & \text{for } \; \Delta_1 < z \le \Delta_2, \\ C_2 z^{k_2} & \text{for } \; z > \Delta_2, \end{cases}$$

where the constants C_1, Δ_1, k_1, C_2, Δ_2, *and* k_2 *are determined from the following system:*

$$(22) \qquad \begin{aligned} &C_1 = (1-\varepsilon)e^{-1+\delta}(1-\delta)^{1-\delta}, \quad \Delta_1 = 1-\delta, \quad k_1 = -1+\delta, \\ &C_2 = (1-\varepsilon)e^{-1-\delta}(1+\delta)^{1+\delta}, \quad \Delta_2 = 1+\delta, \quad k_2 = -1-\delta, \\ &\frac{e^{\delta} + e^{-\delta}}{e\delta} = \frac{1}{1-\varepsilon}. \end{aligned}$$

Proof. Applying the assertion of Theorem 1 to the exponential distribution, we can easily obtain the statement. □

In formula (22), the auxiliary parameter δ $(0 < \delta \le 1)$ is introduced. The expressions for the Fisher information have the following forms for solutions (19) and (21), respectively:

$$I(f^*) = 1 - \varepsilon\Delta^2, \quad I(f^*) = 2\delta\,\tanh(\delta) - \delta^2.$$

With small values of ε, the least favorable density f^* in (19) corresponds to the exponential distribution in the zone $0 < z \le \Delta$; in the "tail" zone, it is similar to the one-sided t-distribution. With large values of ε, the rather whimsical distribution minimizing the Fisher information appears–its density tends to infinity at $z = 0$. The border between these solutions is characterized by the following values of the parameters: $\Delta_1 = 0$, $\Delta_2 = 2$, $\varepsilon = 0.119$. Some numerical results are represented in Table 1. The values of the distribution function

$$F^*(z) = \int_0^z f^*(t)dt,$$

evaluated at the points of "gluing" of the extremals $C_1 z^{k_1}$ and $C_2 z^{k_2}$ with the restriction curve $(1 - \varepsilon)e^{-z}$ are also given in Table 1. The asymptotically efficient estimator $\widehat{T}_n$, evaluated from equation (2), has the following score function

$$\chi^*(z) = -z \, \frac{(f^*(z))'}{f^*(z)} - 1 = \begin{cases} \Delta_1 - 1, & 0 < z \le \Delta_1, \\ z - 1, & \Delta_1 < z \le \Delta_2, \\ \Delta_2 - 1, & z > \Delta_2. \end{cases} \tag{23}$$

Formula (23) is valid for the both solutions (19) and (21): solution (21) comes to solution (19) with $\Delta_1 = 0$. With the notations introduced in Section 3, we obtain from (2) and (23)

$$\sum_{i \in I_1} (\Delta_1 - 1) + \sum_{i \in I} \left(\frac{t_i}{\widehat{T}_n} - 1 \right) + \sum_{i \in I_2} (\Delta_2 - 1) = 0 \tag{24}$$

and

$$\widehat{T}_n = \frac{1}{n - n_1 \Delta_1 - n_2 \Delta_2} \sum_{i \in I} t_i. \tag{25}$$

Estimator (25) is similar to the trimmed mean

$$\widehat{T}_n(n_1, n_2) = \frac{1}{n - n_1 \Delta_1 - n_2 \Delta_2} \sum_{i = n_1 + 1}^{n - n_2} t_{(i)}, \tag{26}$$

where $t_{(i)}$ is an i-th order statistic. If the numbers of the trimmed order statistics (the smallest and greatest) are chosen as

$$n_1 = [F^*(\Delta_1)n], \;\; n_2 = [(1 - F^*(\Delta_2))n],$$

where $[\cdot]$ stands for the integer part of a number in the blank, then the estimators $\widehat{T}_n$ obtained from equation (17) and $\widehat{T}_n(n_1, n_2)$ are asymptotically equivalent as M- and L-estimators of scale ([4]). Hence the simple estimator $\widehat{T}_n(n_1, n_2)$ is recommended for the practical use.

ε	Δ_1	Δ_2	$F^*(\Delta_1)$	$F^*(\Delta_2)$	$1/I(f^*)$
0	0	∞	0	1	1
0.001	0	5.42	0	0.995	1.03
0.002	0	4.86	0	0.990	1.05
0.005	0	4.15	0	0.979	1.09
0.01	0	3.63	0	0.964	1.15
0.02	0	3.13	0	0.937	1.24
0.05	0	2.52	0	0.874	1.47
0.1	0	2.10	0	0.790	1.79
$\varepsilon_0 = (1+e^2)^{-1} \approx 0.119$	0	2	0	0.762	1.91
0.15	0.110	1.890	0.094	0.727	2.11
0.20	0.226	1.774	0.187	0.689	2.46
0.25	0.313	1.687	0.249	0.659	2.88
0.30	0.384	1.616	0.297	0.635	3.38
0.40	0.503	1.497	0.367	0.596	4.76
0.50	0.603	1.397	0.415	0.565	7.03
0.65	0.733	1.267	0.462	0.532	14.7
0.80	0.851	1.149	0.462	0.510	45.8
1	1	1	0.5	0.5	∞

TABLE 1. ε-contaminated exponential distributions minimizing the Fisher information for a scale parameter

The structure of the least informative density and the corresponding structure of the score function show that, for small values of the contamination parameter ε, the optimal algorithm provides the one-side sample trimming with subsequent averaging of the remained sample elements. For large values of ε, the two-side trimming of the smallest and greatest observations is realized.

The practical recommendations for the use of the designed robust estimator are defined by the restrictions of model (1) and within the frames of this model by the value of the contamination parameter ε. The results of investigations in various areas of technical and engineering applications of statistical methods show the good fit of the contamination model to the real-life data [3]. The estimated and expected values of ε usually lie in the interval $(0.001, 0.1)$. If there is no any prior information about the value of ε then one may take it equal to 0.1. In this case, according to the results represented in Table 1 (see its eighth row), the optimal estimator is the one-sided trimmed mean at the level 21%. The deletion of the 21% of the greatest time to failure values and averaging of the rest of them gives perhaps not very optimistic but the guaranteed reliable value of the mean time to failure characteristic.

REMARK 2. It is easy to see that in the gross error model (1) with the arbitrary small ε and the Cauchy-type density $h(t)$, the use of the classical sample mean estimator $\overline{T}$ leads to infinite losses in efficiency as compared with the use of the obtained robust estimators.

References

[1] A. Adatia, *Robust estimators of the 2-parameter gamma distribution*, IEEE Trans. Rel. **37** (1988), 234–238.

[2] M. A. Gallagher and A. H. Moore, *Robust minimum distance estimation using the 3-parameter Weibull distribution*, IEEE Trans. Rel. **39** (1990), 575–580.

[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, John Wiley, New York, 1986.

[4] P. J. Huber, *Robust Statistics*, John Wiley, New York, 1981.

[5] ———, *Robust estimation of a location parameter*, Ann. Math. Stat. **35** (1964), 73–101.

[6] T. Seki and S. Yokagama, *Robust parameter-estimation using the bootstrap method for the 2-parameter Weibull*, IEEE Trans. Rel. **45** (1996), 34–41.

Jae Won Lee, Department of Applied Mathematics, Kumoh National University of Technology, Kumi 730–701, Korea
E-mail: ljaewon@knut.kumoh.ac.kr

Georgy L. Shevlyakov, Department of Mathematics, St. Petersburg State Technical University, St. Petersburg 195251, Russia
E-mail: shev@stat.stu.neva.ru