Robust Minimax Estimation of a Scale Parameter of Exponential Distribution

G.L. Shevlyakov

1.Introduction. The exponential distribution is a simplest model for the discription of time to failure in the reliability theory. The effective maximum likelihood estimate of the failure intensity parameter λ is the inverse value of the sample mean of the time to failure data observations: $\hat{\lambda} = 1/\overline{T}$, $\overline{T} = \sum t_i$. The linear structure of the sample mean type estimators results in their marked unstability to the occasional appearance of rare outliers in data. From the statistical point of view this unstability causes a sharp loss of the estimators' efficiency with small deviations from the accepted stochastic model of data distribution which in turn may lead to great errors in designed reliability characteristics.

In mathematical statistics, robust methods are used in order to provide the stability of statistical inference under the uncertainty of distribution models [1]. The main results in robust statistics refer to the normal and its neighbourhood models of data distributions. The goal of this article is to use robust minimax approach to one of the traditional for the reliability theory models of data distributions.

2. Problem statement. Consider a sample $t_1, ..., t_n$ of i. i. d. random variables from the mixture model of distributions [1]

$$\mathcal{F} = \{\mathrm{f} : \mathrm{f}\ (\mathrm{t}, \mathrm{T}) = (1-\varepsilon)\mathrm{T}^{-1}\exp(-\mathrm{t}/\mathrm{T}) + \varepsilon\ \mathrm{h}(\mathrm{t}),\ \ 0 \le \varepsilon < 1\}, \tag{1}$$

where ε is a contamination parameter characterizing the level of uncertainty of the accepted exponential model; $h(t)$ is an arbitrary p. d. f.; T is a scale parameter under estimation.

Following Huber [1], consider M - estimates of a scale parameter T

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \chi\left(\frac{\mathrm{t_i}}{\hat{\mathrm{T}}_\mathrm{n}}\right) = 0\ , \tag{2}$$

where χ is a defining function for the class (2).

In the important particular case with the known p. d. f. $T^{-1}f(t/T)$ the following choice of the defining function

$$\chi(\mathrm{z}) = -\mathrm{z}\,\frac{\mathrm{f}'(\mathrm{z})}{\mathrm{f}\ (\mathrm{z})} - 1 \tag{3}$$

gives the maximum likelihood estimates as the solutions of the equation (2).

Under some definite restrictions of regularity put on the class of p. d. f. $\mathcal{F}$ (1) and on the defining functions $\chi(z)$ (later they will be formulated), and with the condition

$$\int_0^\infty \chi(\mathrm{z})\mathrm{f}(\mathrm{z})\,\mathrm{dz} = 0 \tag{4}$$

the estimates $\hat{T}_n$ are consistent, asymptotically normal and asymptotically efficient, and possess the minimax property with respect to the asymptotic variance $D\hat{T}_n = V(\chi, f)$

$$\mathrm{V}(\chi^*, \mathrm{f}) \le \mathrm{V}(\chi^*, \mathrm{f}^*) \le \mathrm{V}(\chi, \mathrm{f}^*), \tag{5}$$

where f^* is a least favorable p. d. f., minimizing the Fisher information $I(f)$ for a scale parameter on the class (1)

$$\mathrm{f}^* = \arg\min \mathrm{I}(\mathrm{f}), \qquad \mathrm{I}(\mathrm{f}) = \int_0^\infty \mathrm{z}^2 \left(\frac{\mathrm{f}'(\mathrm{z})}{\mathrm{f}(\mathrm{z})}\right)^2 \mathrm{f}(\mathrm{z})\,\mathrm{dz} - 1; \tag{6}$$

$\chi^*(z)$ is determined from the equation (3) [1].

The condition (5) defines the saddle-point (χ^*, f^*) of the asymptotic variance $V(\chi, f)$. The left part of the inequality is of a practical importance: the choice of the estimate (2) with the defining function $\chi^*(z)$ provides the guaranteed level of the estimate's accuracy for all p. d. f. belonging to the class (1):

$$\mathrm{V}(\chi^*, \mathrm{f}) \le \mathrm{V}(\chi^*, \mathrm{f}^*) = \frac{1}{\mathrm{n}\,\mathrm{I}(\mathrm{f}^*)} \tag{7}$$

3. The least favorable density f^* . The problem of the design of the minimax estimate $\hat{T}_n$ comes formally to the variational problem (6) with the additional conditions of nonnegativeness and norming

$$\mathrm{f}^* = \arg\min \mathrm{I}(\mathrm{f}), \tag{8}$$

$$\int_0^\infty \mathrm{f}(\mathrm{z})\,\mathrm{dz} = 1, \tag{9}$$

$$\mathrm{f}(\mathrm{z}) \ge (1-\varepsilon)\,\mathrm{e}^{-\mathrm{z}}, \tag{10}$$

where $T = 1$ is taken for convinience. The restriction of the class (1) is written in the form (10), respectively it includes the nonnegativeness condition.

Theorem. *The solution of the problem (8) with the restrictions (9) and (10) for the class of continuously- differentiable on* $(0, \infty)$ *p. d. f.* $f(z)$ *is of the following form:*

with $0 \le \varepsilon < \varepsilon_0 = (1+e^2)^{-1}$

$$f^*(z) = \begin{cases} (1-\varepsilon)\,e^{-z}, & \text{with } 0 \le z < \Delta, \\ C\,z^k, & \text{with } z \ge \Delta, \end{cases} \tag{11}$$

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

$$C = (1-\varepsilon)\, e^{-\Delta}\, \Delta^{\Delta}, \;\; k = -\Delta, \;\; \frac{e^{-\Delta}}{\Delta - 1} = \frac{\varepsilon}{1-\varepsilon}\;; \tag{12}$$

with $\varepsilon_0 \le \varepsilon < 1$

$$f^*(z) = \begin{cases} C_1 z^{k_1}, & with\ 0 \le z < \Delta_1, \\ (1-\varepsilon)\, e^{-z}, & with\ \Delta_1 \le z < \Delta_2, \\ C_2 z^{k_2}, & with\ \ z \ge \Delta, \end{cases} \tag{13}$$

where the constants $C_1, \Delta_1, k_1, C_2, \Delta_2, k_2$ *are determined from the following equations*

$$\begin{gathered} C_1 = (1-\varepsilon) e^{-1+\delta} (1-\delta)^{1-\delta}, \;\Delta_1 = 1-\delta, \; k_1 = -1+\delta, \\ C_2 = (1-\varepsilon) e^{-1-\delta} (1+\delta^{1+\delta}), \;\Delta_2 = 1+\delta, \; k_2 = -1-\delta, \\ \frac{e^{\delta} + e^{-\delta}}{e\delta} = \frac{1}{1-\varepsilon}\;. \end{gathered} \tag{14}$$

The proof of the theorem is in the Appendix.

In the formulae (14) the auxiliary parameter $\delta\ (0 \le \delta \le 1)$ is introduced. The expressions for the Fisher information are of the following form, respectively for the solutions (11) and (13):

$$\mathrm{I(f^*)} = 1 - \varepsilon\Delta^2, \;\; \mathrm{I(f^*)} = \frac{2\delta}{\mathrm{th}\delta} - \delta^2. \tag{15}$$

With the small values of ε the least favourable density f^* (11) corresponds to the exponential distribution in the zone $0 \le z < \Delta$; in the "tail" zone it is similar to the one-sided t - distribution. With the 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 Tab. 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 $(1-\varepsilon)\, e^{-z}$ are also given in Tab. 1.

4. The structure of the robust minimax estimate. The asymptotically effective estimate $\hat{T}_n$, evaluated from the equation (2), has the following defining function $\chi(z)$

$$\chi^*(\mathrm{z}) = -\mathrm{z}\,\frac{(\mathrm{f^*(z)})'}{\mathrm{f^*(z)}} - 1 = \begin{cases} \Delta_1 - 1, & 0 \le z < \Delta_1, \\ z - 1, & \Delta_1 \le z < \Delta_2, \\ \Delta_2 - 1, & z \ge \Delta_2. \end{cases} \tag{16}$$

The formula (16) is valid for the both solutions (11) and (13): the solution (13) comes to the solution (11) with $\Delta_1 = 0$.

Consider the following notations:

$$I_1 = \{i : t_i/\hat{T}_n < \Delta_1\},\ I_2 = \{i : t_i/\hat{T}_n \geq \Delta_2\},\ I = \{i : \Delta_1 \leq t_i/\hat{T}_n < \Delta_2\}.$$

The equation (2) is written then as

$$\sum_{\mathrm{i \in I_1}} (\Delta_1 - 1) + \sum_{\mathrm{i \in I}} (\frac{\mathrm{t_i}}{\hat{\mathrm{T}}_\mathrm{n}} - 1) + \sum_{\mathrm{i \in I_2}} (\Delta_2 - 1) = 0\ . \tag{17}$$

Denoting the numbers of observations belonging to the sets I_1, I_2 and I respectively as n_1, n_2 and $n - n_1 - n_2$, we get from (17)

$$\hat{\mathrm{T}}_\mathrm{n} = \frac{1}{\mathrm{n} - \mathrm{n_1}\Delta_1 - \mathrm{n_2}\Delta_2} \sum_{\mathrm{i \in I}} \mathrm{t_i}\ . \tag{18}$$

The structure of the estimate (18) is similar to the structure of the trimmed mean

$$\hat{\mathrm{T}}_\mathrm{n}(\mathrm{n_1}, \mathrm{n_2}) = \frac{1}{\mathrm{n} - \mathrm{n_1}\Delta_1 - \mathrm{n_2}\Delta_2} \sum_{\mathrm{i = n_1 + 1}}^{\mathrm{n - n_2}} \mathrm{t_{(i)}}\ , \tag{19}$$

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

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

where $[\cdot]$ is an integer part of a number then the estimates $\hat{T}_n$ obtained from the equation (17) and $\hat{T}_n(n_1, n_2)$ are asyptotically equivalent. So, the simple estimate $\hat{T}_n(n_1, n_2)$ (19) is recommended for the practical use.

Notice that in the limit case with $\varepsilon \to 0$, the robust minimax estimate defined by the numerical solution of the equation (17) (or by the formula (19)) is the sample median

$$\hat{\mathrm{T}}_\mathrm{n} = \mathrm{med}\ \mathrm{t_i} = \begin{cases} t_{(h+1)}, & n = 2h + 1, \\ (t_{(h)} + t_{(h+1)})/2, & n = 2h. \end{cases}$$

With $\varepsilon = 0$ the robust minimax estimate $\hat{T}_n$ is the sample mean $\overline{T}$ - the effective estimate of a scale parameter of the exponential distribution.

5. Conclusions. The structure of the least favorable density and the corresponding structure of the defining function show that with the small values of the contamination parameter ε the optimum algorithm provides the one-side sample trimming with the following averaging of the remaining sample elements. With the large values of ε the two-side trimming of the least and the largest sample elements is realized.

The practical recommendations for the use of the designed robust estimator (19) are defined by the restrictions of the contamination model (1) and within the frames of this model by the value

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	4.16	0	0.979	1.09
0.01	0	3.63	0	0.964	1.15
0.02	0	3.15	0	0.938	1.25
0.05	0	2.52	0	0.874	1.47
0.1	0	2.11	0	0.791	1.80
0.119	0	2	0	0.762	1.91
0.15	0.110	1.890	0.094	0.727	2.12
0.20	0.225	1.775	0.185	0.679	2.47
0.25	0.313	1.687	0.250	0.659	2.88
0.30	0.384	1.616	0.297	0.635	3.39
0.40	0.503	1.497	0.367	0.595	4.81
0.50	0.603	1.397	0.416	0.565	7.03
0.65	0.733	1.267	0.462	0.532	14.7
0.80	0.851	1.149	0.488	0.511	45.9
1	1	1	0.5	0.5	∞

Table 1: ε - contaminated exponential distributions minimizing the Fisher information for a scale parameter

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 with data [1]. The estimated and expected values of ε as usual are in the interval $(0.001, 0.1)$. If there is no a prior information about the value of ε then one may put it equal to 0.1. In this case according to the formula (19) and to the results represented in Tab.1 the estimator is the one-sided trimmed mean at the level 21 %.

The rejection of the 21 % of the largest time to failure values and the averaging of the remaining gives perhaps not very optimistic but the guaranteed reliable value of the mean time to failure characteristic.

APPENDIX

First we obtain the structure of the solutions (11) and (13), and then prove their optimality.

The variational problem (8) with the restriction (9) is reformulated by the use of the following change of variables $f(z) = g^2(z) \geq 0$

$$J(g) = \int_0^\infty z^2 g'(z)^2 \, dz \to \min \,, \qquad \int_0^\infty g^2(z) \, dz = 1. \qquad (A1)$$

The Lagrange functional for the problem (A1) is

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

The Euler equation for it is of the form

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

and respectively its solutions are the extremals of the problem (A1)

$$f_1(z) = g_1^2(z) = C_1 z^{k_1}, \;\; f_2(z) = g_2^2(z) = C_2 z^{k_2}, \;\; k_1 + k_2 = -2 \qquad (A2)$$

The optimum solution of the original problem (8) with the restictions (9) and (10) is the smooth "gluing" of the free extremals (A2) and the restriction $(1-\varepsilon)\, e^{-z}$ (10) in the form of (11) and (13). The parameters of "gluing" $C_1, \Delta_1, k_1, C_2, \Delta_2, k_2$ are determined from the conditions of norming, continuity and differentiability of the solution at the points $z = \Delta_1$ and $z = \Delta_2$ what gives the equations (12) and (14).

It is known [1], that the density f^* belonging to the class of convex densities $\mathcal{F}$ (1) minimizes the Fisher information iff

$$[\frac{d}{dt} I(f_t)]_{t=0} \geq 0 \,, \qquad (A3)$$

where $f_t = (1-t)f^* + tf$, and f is an arbitrary p. d. f. with $I(f) < \infty$.

The inequality (A3) is written as

$$\int_0^\infty (2\chi^{*\prime} - \chi^{*2})(f - f^*) \, dt \geq 0 \,, \qquad (A4)$$

where $\chi^*(z)$ is the defining function (16).

The direct evaluation of the left part of (A4) gives

$$f(t) - f^*(t) = f(t) - (1-\varepsilon)\, e^{-t} \geq 0$$

so the inequality (A3) is equivalent to the restriction (10) of the variational problem, and this remark finishes the proof.

REFERENCES

1. Huber, P.J. (1981) Robust Statistics. Wiley, New York.