On the Characterization of Fisher Information and Stability of the Least Favourable Lattice Distributions

VIL'CHEVSKIY NIKITA O. and GEORGIY L. SHEVLYAKOV

(Department of Mathematics, St.Petersburg State Technical University,
Polytechnicheskaya st., 29, St.Petersburg, 195251, Russia)

ABSTRACT: *The paper is concerned with the stability properties of the least favourable distributions minimizing Fisher information in a given class of distributions. The derivation of a least favourable distribution (the solution of a variational problem) is a necessary stage of the Huber's minimax approach in robust estimation of a location parameter. Generally, the solutions of variational problems essentially depend on the regularity restrictions of a functional class.*

The stability of these optimal solutions to the violations of smoothness restrictions is studied under the lattice distribution classes. The discrete analogues of Fisher information are obtained in these cases. They have the form of the Hellinger metrics with the estimation of a real continuous location parameter and and the form of the χ^2 metrics with the estimation of an integer discrete location parameter.

The analytical expressions for the corresponding least favourable discrete distributions are derived in some classes of lattice distributions by means of generating functions and Bellman's recursive functional equations of the dynamic programming. These classes include the class of nondegenerate distributions with the restriction on the value of a density in the centre of symmetry, the class of finite distributions and the class of contaminated distributions.

The obtained least favourable lattice distributions preserve the structure of their prototypes in the continuous case. These results show the stability of robust minimax solutions under the different types of transitions from the continuous distribution to the discrete one.

1. Introduction

One of the basic approaches to the synthesis of robust estimation procedures is the minimax principle. In this case, under a given class of distributions, the least favourable distribution (which minimizes the Fisher information) is determined. The unknown parameters of a distribution model are then estimated by means of the maximum likelihood method for this distribution [1].

As a result, if we know the set of possible deviations of the actual probability distribution from the model that we are currently using, then we can construct *robust* statistical procedures, i.e., procedures which are stable with respect to (possible) deviations from an apriori distribution model.

The robust minimax procedures provide a guaranteed level of the estimator's accuracy (measured by the supremum of an asymptotic variance) for any distribution of a given class.

The form of the solution obtained by the minimax approach essentially depends upon the characteristics of a distribution class. As a rule, the classes of continuous and symmetric distributions are considered [1]. In many real-life problems of data processing, the results of measurements include groops of equal values. Furthermore, the results of measurements usually come rounded in accordance with the scale of a measuring device playing a role of a discretizator. Thus, in this cases, the use of continuous distribution models seems not adequate to the original problem of data processing, and it is rather important for applications to design robust methods for discrete distribution models that correspond to the real nature of data.

In this paper, we describe the analogues of Fisher information for the discrete distribution classes considering:

- the direct disretization procedure of the Fisher information functional in the problem estimation of a continuous location parameter;
- the discrete analogue of the Rao-Cramer inequality in the problem of estimation of a discrete location parameter.

In the last case, the obtained form of the Rao-Cramer inequality is similar to the Chapman-Robbins inequality [2].

The derived terms corresponding to the Fisher information functional are different in the considered cases but the solutions of the variational problems of minimization of these functionals (the least favourable distributions) are the same. Moreover, they demonstrate a remarkable correspondence with their continuous analogues. Thus, we can conclude that the structure of robust minimax procedures [1] is rather stable to the deviations from the assumptions of regularity of distribution classes.

2. The Least Favourable Distributions in the Continuous Case

The main stage of the minimax approach to the design of a robust minimax estimate of a location parameter θ of distribution density $f(x,\theta) = f(x-\theta)$ belonging to a certain class $\mathcal{F}$ is the solution of the variational problem of the minimization of the Fisher information [1]

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

For all considered classes of distributions $\mathcal{F}$, the following conditions are common:

$$f(x) \geq 0, \quad f(-x) = f(x), \quad \int_{-\infty}^{\infty} f(x)dx = 1. \tag{2}$$

Depending on the additional restrictions put upon the class $\mathcal{F}$, different forms of the density f^* may result.

Consider the following basic classes of distributions in robust estimation:

- the class of nonsingular densities

$$\mathcal{F}_1 = \left\{ f : f(0) \geq \frac{1}{2a} > 0 \right\}$$

with the Laplace least favourable density

$$f_1^*(x) = \frac{1}{2a} \exp(-|x|/a);$$

- the class of ε - contaminated distributions

$$\mathcal{F}_2 = \{f : f(x) \geq (1-\varepsilon)p(x), \;\; 0 < \varepsilon < 1\}$$

with the least favourable density having the exponential "tails"

$$f_2^*(x) = \begin{cases} (1-\varepsilon)p(x), & |\, x \,| \leq \Delta, \\ Af_1^*(Bx), & |\, x \,| > \Delta, \end{cases}$$

where $p(x)$ is a given density; ε is a parameter characterising the degree of apriori uncertainty; the constants A, B, Δ are chosen to satisfy the conditions of normalization and the sewing smoothness at the point $x = \Delta$;
- the class of finite distributions

$$\mathcal{F}_3 = \left\{ f : \int_{-a}^{a} f(x)\, dx = 1 \right\}$$

with the following least favourable density

$$f_3^*(x) = \begin{cases} a^{-1} \cos^2((\pi x)/(2a)), & |\, x \,| \leq a, \\ 0, & |\, x \,| > a. \end{cases}$$

The essential characteristic feature of these results presented in [1] is that the least favourable densities are of the "exponential" type (the cos-type density f_3^* is too), and it is caused by the structure of the extremals of the variational problem (1) [3]. Hence, the least favourable exponential tails imply the maximum likelihood procedures with rejection of outliers [1].

3. The Discrete Analogues of Fisher Information

Consider the class of lattice distributions:

$$f_l(x) = \sum_i p_i \delta(x - i\Delta), \quad \sum_i p_i = 1, \tag{3}$$

where $\delta(\cdot)$ is a delta-function of Dirac, Δ is a step of discretization.
We consider two different cases :
- a location parameter is continuous with $\theta \in \mathbf{R}$;
- a location parameter is discrete with $\theta \in \mathbf{Z}$.

In the first case the following result is valid:

Theorem 1 *Under the class of lattice distributions (3) with a continuous parameter θ the variational problem (1) is equivalent to the following variational problem*

$$\sum(\sqrt{p_{i+1}} - \sqrt{p_i})^2 \to \min \ . \tag{4}$$

The proofs are carried out into Appendix.

In the second case with a discrete location parameter θ the following analogue of the Rao-Cramer inequalty is valid:

Theorem 2 *Let $x_1, ..., x_n$ be i.i.d.r.v. with the distribution density $f(x - \theta)$ (3) and $p_i > 0, i \in \mathbf{Z}$; $x_1, ..., x_n, \theta \in \mathbf{Z}$;*
Let $\hat{\theta}_n = \hat{\theta}_n(x_1, ..., x_n)$ be a discrete unbiased estimate of a discrete location parameter:

$$\hat{\theta}_n \in \mathbf{Z}, \quad \mathbf{E}\hat{\theta}_n = \theta.$$

Then the following inequality holds for the variance of this estimate:

$$\mathbf{D}\hat{\theta}_n \geq \frac{1}{\left(\sum\limits_{i\in\mathbf{Z}} \frac{(p_{i-1}-p_i)^2}{p_i} + 1\right)^n - 1}. \tag{5}$$

The essential feature of the obtained result is that, in the discrete case, the lower boundary of the estimate's variance decreases exponentially with $n \to \infty$ providing the corresponding efficiency of estimation much greater than in the continuous case.

Corollary 1 *Under the class of lattice distributions (3) with a discrete parameter θ, the variational problem (1) is equivalent to the following variational problem*

$$\sum_{i\in\mathbf{Z}} \frac{p_{i-1}^2}{p_i} \to \min \ . \tag{6}$$

4. The Discrete Least Favourable Distributions

In this section, we consider the discrete analogues of the least favourable distributions for the continuous classes $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$ enlisted in Section 2.

Let $\mathcal{P}_1$ be a class of lattice symmetric nonsingular distributions

$$\mathcal{P}_1 = \Big\{p_i, \ i \in \mathbf{Z} : p_i > 0, \ \ p_0 \geq \gamma_0 > 0, \ \ p_{-i} = p_i, \ \ \sum p_i = 1\Big\}.$$

Theorem 3 *Under the class of lattice distributions $\mathcal{P}_1$, the solution of the variational problem (4) is of the form:*

$$p^*_{-i} = p^*_i = \alpha^i\, \gamma_0, \;\; \alpha = \frac{1-\gamma_0}{1+\gamma_0}, \;\; i = 0, 1, \dots\, . \tag{7}$$

Theorem 4 *Under the class of lattice distributions $\mathcal{P}_1$, the solution of the variational problem (6) is the same as in Theorem 3.*

The least favourable lattice distribution f^*_{l1} (3) with the geometric progression of p^*_i is the discrete analogue of the least favourable Laplace density f^*_1 for the continuous distribution class $\mathcal{F}_1$.

Consider the discrete analogue of the class of ε - contaminated distributions $\mathcal{F}_2$ with the restrictions upon the values of p_i in the central zone of a distribution:

$$\mathcal{P}_2 = \left\{ p_i,\; i \in \mathbf{Z} : p_i > 0, \quad p_{-i} = p_i \geq \gamma_i > 0, \;\; i = 0, 1, ..., k; \;\; \sum p_i = 1 \right\}.$$

Theorem 5 *Under the class of lattice distributions $\mathcal{P}_2$ with the additional resrictions put upon the given values of γ_i*

$$\gamma_i^{1/2} - \gamma_{i-1}^{1/2} \leq \frac{(1-\alpha^{1/2})^2}{2\alpha^{1/2}} \sum_{j=0}^{i-1} \gamma_j^{1/2},$$

the solution of the variational problem (4) is of the form:

$$p^*_{-i} = p^*_i = \begin{cases} \gamma_i, & i = 0, 1, ..., s^*, \;\; s^* \leq k, \\ \alpha^{i-s^*}\gamma_{s^*}, & i > s^*, \end{cases} \tag{8}$$

where

$$\alpha = (1 - \gamma_0 - 2\sum_{i=0}^{s^*} \gamma_i)/(1 - \gamma_0 - 2\sum_{i=0}^{s^*} \gamma_i + 2\gamma_{s^*})\;;$$

the sewing number s^* is determined by the maximum value of s satisfying the following restrictions:

$$2(\gamma_{s-1}\gamma_s)^{1/2} + \left(1 - \gamma_0 - 2\sum_{i=0}^{s-1}\gamma_i\right)^{1/2} \left(\left(1 - \gamma_0 - 2\sum_{i=0}^{s}\gamma_i\right)^{1/2} - \left(1 - \gamma_0 - 2\sum_{i=0}^{s-2}\gamma_i\right)^{1/2} \right) > 0.$$

The connection of this result with the Huber's least favourable density f^*_2 is obvious.

Finally, consider the discrete analogue of the class of finite distributions $\mathcal{F}_3$:

$$\mathcal{P}_3 = \left\{ p_i,\; i \in \mathbf{Z} : p_i \geq 0, \quad p_{-i} = p_i > 0, \;\; i = 0, 1, ..., n; \;\; p_i = 0, \;\; i > n; \;\; \sum p_i = 1 \right\}.$$

Theorem 6 *Under the class of lattice distributions $\mathcal{P}_3$ the solution of the variational problem (4) is of the form:*

$$p^*_{-i} = p^*_i = \frac{1}{n+1}\cos^2\left(\frac{i\pi}{2(n+1)}\right), \quad i = 0, ..., n. \tag{9}$$

The results of Theorems 3, 4, 5 and 6 show the stability of robust minimax solutions under the violations of regularity conditions of distribution classes caused by different types of transitions from the continuous to the discrete case.

Appendix

Proof of Theorem 1. In the variational problem (1), the condition of nonnegativeness of a density $f \geq 0$ is accounted with the following change of variables $f = g^2$:

$$I(g) = \int_{-\infty}^{\infty} (g'(x))^2 \, dx \; \to \min_g, \quad \int_{-\infty}^{\infty} g^2(x) \, dx = 1. \tag{10}$$

Consider the δ - sequence approximation of the formula (3) with $\Delta = 1$:

$$f_h(x) = g_h^2(x), \;\; g_h(x) = \sum_i \frac{p_i^{1/2}}{2\pi h^2} \exp\left\{-\frac{(x-i)^2}{4h^2}\right\}.$$

In this case, the functional (10) and the norming condition are written:

$$I_h = \frac{1}{h^2} - \frac{1}{4h^4}\sum_i\sum_j \sqrt{p_i}\sqrt{p_j}(i-j)^2 \exp\left\{-\frac{(i-j)^2}{8h^2}\right\}, \quad \sum_i\sum_j \sqrt{p_i}\sqrt{p_j} \exp\left\{-\frac{(i-j)^2}{8h^2}\right\} = 1.$$

The main part of the functional I_h with $h \to 0$ is $-\sum \sqrt{p_i}\sqrt{p_{i+1}}$. Taking into account the norming condition for p_i, we get the statement of Theorem 1.

Proof of Theorem 2. Consider the likelihood

$$L(x_1, ..., x_n|\theta) = p_{x_1-\theta} \cdots p_{x_n-\theta}.$$

The norming and unbiasedness conditions are in this case:

$$\sum_{x_1,...,x_n \in \mathbf{Z}} L(x_1, ..., x_n|\theta) = 1, \tag{11}$$

$$\sum_{x_1,...,x_n \in \mathbf{Z}} \hat{\theta}_n(x_1, ..., x_n) L(x_1, ..., x_n|\theta) = \theta. \tag{12}$$

Writing out the unbiasedness condition for the parameter value $\theta+1$

$$\sum_{x_1,...,x_n\in\mathbf{Z}} \hat{\theta}_n(x_1,...,x_n)L(x_1,...,x_n|\theta+1)=\theta+1$$

and substracting the expression (12) from it, we have

$$\sum_{x_1,...,x_n\in\mathbf{Z}} \hat{\theta}_n(x_1,...,x_n)\left[L(x_1,...,x_n|\theta+1)-L(x_1,...,x_n|\theta)\right]=1. \tag{13}$$

Denote $\hat{\theta}_n(x_1,...,x_n)=\hat{\theta}_n$ and $L(x_1,...,x_n|\theta)=L(\theta)$. Then with the norming condition (11) we get from (13)

$$\sum_{x_1,...,x_n\in\mathbf{Z}} (\hat{\theta}_n-\theta)\left[\frac{L(\theta+1)-L(\theta)}{L(\theta)}\right]L(\theta)=1. \tag{14}$$

Finally, the Cauchy-Bunyakovskiy inequality and the formula (14) imply

$$\sum_{x_1,...,x_n\in\mathbf{Z}} (\hat{\theta}_n-\theta)^2L(\theta) \sum_{x_1,...,x_n\in\mathbf{Z}} \left[\frac{L(\theta+1)-L(\theta)}{L(\theta)}\right]^2 L(\theta)\geq 1$$

and the Rao-Cramer type inequality in the following form

$$\mathbf{D}\hat{\theta}_n \geq \frac{1}{\sum\limits_{x_1,...,x_n\in\mathbf{Z}} \left[\frac{L(\theta+1)-L(\theta)}{L(\theta)}\right]^2 L(\theta)}. \tag{15}$$

The statement of Theorem 2 is directly obtained from the formula (15).

Proof of Theorem 3. Denote $\lambda_i=\sqrt{p_i},\;\; i\in\mathbf{Z}$. Let the parameter $\lambda_0=\sqrt{p_0}\geq\sqrt{\gamma_0}>0$ be free, the optimization by it will be done at a final stage of a solution.

The variational problem (4) may be reformulated as the following:

$$\sum_{i\in\mathbf{Z}} \lambda_i\lambda_{i+1}\to\max_{\Lambda}, \tag{16}$$

where $\Lambda=\{\lambda_i,\;\; i\in\mathbf{Z}\}$.

In this case, the Lagrange functional is of the form:

$$2(\sqrt{p_0}\lambda_1+\sum_{i=1}^{\infty}\lambda_i\lambda_{i+1})-\mu(p_0+2\sum_{i=1}^{\infty}\lambda_i^2-1)\to\max_{\mu,\Lambda}, \tag{17}$$

where μ is a Lagrangian multiplier corresponding to the norming condition.

The extremum conditions for the problem (17) are given by the following infinite system of equations:

$$\begin{cases} \sqrt{p}_0 \ - \ 2\mu\lambda_1 + \lambda_2 = \ 0, \\ \lambda_1 \quad - \ 2\mu\lambda_2 \ + \lambda_3 = \ 0, \\ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ , \\ \lambda_{k-1} - 2\mu\lambda_k + \lambda_{k+1} = 0, \\ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \end{cases} \tag{18}$$

For the solution of the system (18), let introduce a generating function

$$F(x) = \sum_{i=0}^{\infty} \lambda_{i+1} x^i, \quad |x| < 1. \tag{19}$$

We obtain the evident expression for (19) by multiplying the equations (18) by x^i $(i = 0, 1, ...,)$ and summarizing them:

$$F(x) = \frac{\lambda_1 - \sqrt{p}_0 x}{x^2 - 2\mu x + 1}. \tag{20}$$

Set $\lambda_1 = t\sqrt{p}_0$ in (20)

$$F(x) = \frac{t - x}{x^2 - 2\mu x + 1} \sqrt{p}_0. \tag{21}$$

The denominator of (21) expands into the multipliers

$$x^2 - 2\mu x + 1 = (x - x_0)(x - 1/x_0), \ \ x_0 = \mu - \sqrt{\mu^2 - 1},$$

with $x_0 = t$. Hence, the formula (21) takes the form

$$F(x) = \frac{t}{1 - tx} \sqrt{p_0} = t\sqrt{p_0} \sum_{i=0}^{\infty} t^i x^i. \tag{22}$$

Comparing the series (19) and (22) we get

$$\lambda_i = t^i \sqrt{p_0}, \quad i \in \mathbf{N}.$$

The value of t is determined from the norming condition

$$p_0 + 2p_0 \sum_{i=1}^{\infty} t^{2i} = 1,$$

which gives

$$t = \frac{1 - p_0}{1 + p_0}.$$

The functional (16) depends of the free parameter p_0 as following:

$$2(\sqrt{p_0}\lambda_1 + \sum_{i=1}^{\infty} \lambda_i \lambda_{i+1}) = \sqrt{1 - p_0^2}.$$

With the condition $p_0 \geq \gamma_0 > 0$, we obtain the optimum solution

$$p_0^* = \arg \max_{p_0 \geq \gamma_0 > 0} \sqrt{1 - p_0^2} = \gamma_0.$$

Denoting $\alpha = t^2(p_0^*) = (1 - \gamma_0)/(1 + \gamma_0)$, we get the statement of Theorem 3.

Remark 1 *If the parameter λ_0 is not free with the optimization provided by it, then the following equation is to be added to the system (18): $-\mu\lambda_0 + \lambda_1 = 0$. In our case, it is fulfilled as a strict inequality: $-\mu\sqrt{\gamma_0} + \lambda_1 < 0$.*

Proof of Theorem 4. In the symmetric case $p_{-i} = p_i$ with the free parameter $p_0 \geq \gamma_0 > 0$, the optimization problem (6) is written as:

$$I = \min_{p_1,\dots} \left[\sum_{i=0}^{\infty} \left(\frac{p_i^2}{p_{i+1}} + \frac{p_{i+1}^2}{p_i} \right), \quad \sum_{i=1}^{\infty} p_i = \frac{1 - p_0}{2} \right] - 1. \tag{23}$$

Consider the following auxiliary optimization problem:

$$\sum_{i=0}^{\infty} \left(\frac{p_i^2}{p_{i+1}} + \frac{p_{i+1}^2}{p_i} \right) \to \min_{p_1,\dots}, \quad \sum_{i=1}^{\infty} p_i = b.$$

Denote the optimum value of a functional in this case as

$$\Phi(p_0, b) = \min_{p_1,\dots} \left[\frac{p_0^2}{p_1} + \frac{p_1^2}{p_0} + \sum_{i=1}^{\infty} \left(\frac{p_i^2}{p_{i+1}} + \frac{p_{i+1}^2}{p_i} \right), \quad \sum_{i=1}^{\infty} p_i = b, \;\; p_i \geq 0 \right]$$

or

$$\min_{0 \leq p_1 \leq b} \left[\frac{p_0^2}{p_1} + \frac{p_1^2}{p_0} + \min_{p_2,\dots} \left[\frac{p_1^2}{p_2} + \frac{p_2^2}{p_1} + \sum_{i=2}^{\infty} \left(\frac{p_i^2}{p_{i+1}} + \frac{p_{i+1}^2}{p_i} \right), \quad \sum_{i=2}^{\infty} p_i = b - p_1, \;\; p_i \geq 0 \right] \right] =$$

$$= \min_{0 \leq p_1 \leq b} \left[\frac{p_0^2}{p_1} + \frac{p_1^2}{p_0} + \Phi(p_1, b - p_1) \right].$$

Consider the function:

$$\psi(y) = \min_{z_1,\dots} \left[\frac{y^2}{z_1} + \frac{z_1^2}{y} + \sum_{i=1}^{\infty} \left(\frac{z_i^2}{z_{i+1}} + \frac{z_{i+1}^2}{z_i} \right), \quad \sum_{i=1}^{\infty} z_i = 1, \;\; z_i \geq 0 \right].$$

The following equation is valid:

$$\Phi(p_0, b) = b\psi\left(\frac{p_0}{b}\right).$$

Hence, we get the recursive Bellman equations:

$$\Phi(p_0, b) = \min_{0 \leq p_1 \leq b} \left[\frac{p_0^2}{p_1} + \frac{p_1^2}{p_0} + \Phi(p_1, b - p_1) \right]$$

or

$$b\psi\left(\frac{p_0}{b}\right) = b\min_{0\le z\le 1}\left[\frac{p_0^2}{b^2}\frac{1}{z} + \frac{z^2}{p_0/b} + (1-z)\psi\left(\frac{z}{1-z}\right)\right].$$

The Bellman function $\psi(y)$ satisfies the following functional equation:

$$\psi(y) = \min_{0\le z\le 1}\left[\frac{y^2}{z} + \frac{z^2}{y} + (1-z)\psi\left(\frac{z}{1-z}\right)\right]. \tag{24}$$

It can be directly checked that the solution of (24) is

$$\psi(y) = \frac{1}{1+y} + (1+y)^2. \tag{25}$$

Thus, we have

$$\min_{0\le z\le 1}\left[\frac{y^2}{z} + \frac{z^2}{y} + (1-z)\left[(1-z) + \frac{1}{(1-z)^2}\right]\right] = \min_{0\le z\le 1}\left[\frac{y^2}{z} + \frac{z^2}{y} + (1-z)^2 + \frac{1}{(1-z)}\right].$$

Taking a derivative, we get

$$-\frac{y^2}{z^2} + 2\frac{z}{y} - 2(1-z) + \frac{1}{(1-z)^2} = (z - y(1-z))\left[\frac{2}{y} + \frac{z + y(1-z)}{(1-z)^2z^2}\right] = 0.$$

The derivative equals zero with $z = y/(1+y)$, and that implies (25).

It follows from (23) that the Fisher information is of the form:

$$I = \frac{1-p_0}{2}\psi\left(2\frac{p_0}{1-p_0}\right) - 1 = \frac{4p_0^2}{1-p_0^2}$$

and

$$\min_{p_0\ge\gamma_0>0} I = \frac{4\gamma_0^2}{1-\gamma_0^2}$$

with

$$p_i = \left(\frac{1-\gamma_0}{1+\gamma_0}\right)^i\gamma_0,$$

and this concludes the proof.

Proof of Theorem 5 is based on:

- the solution of the infinite system of equations as (18) with the first equation:

$$\lambda_{s^*+1} - 2\mu\lambda_{s^*+2} + \lambda_{s^*+3} = 0;$$

- the following maximization of a functional (16) with checking the restrictions of the problem

$$p_i \ge \gamma_i > 0, \quad i = 0, 1, ..., k$$

• and the inequalities of a gradient type:

$$\lambda_k - 2\mu\lambda_{k+1} + \lambda_{k+2} < 0 \quad with \;\; 0 \le k \le s^*.$$

Proof of Theorem 6. In this case, the Lagrange functional is of the form:

$$2\sum_{i=0}^{n-1} \lambda_i\lambda_{i+1} - \mu(\lambda_0^2 + 2\sum_{i=1}^{n} \lambda_i^2 - 1) \to \max_{\mu,\lambda_0,\ldots,\lambda_n}, \tag{26}$$

where μ is a Lagrangian multiplier corresponding to the norming condition. The extremum conditions for the problem (26) are given by the following system of equations:

$$\begin{cases} \qquad\quad - \mu\lambda_0 + \lambda_1 = 0, \\ \lambda_0 \;\; - \;\; 2\mu\lambda_1 \;\; + \lambda_2 = 0, \\ . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; . \;\; , \\ \lambda_{n-2} - 2\mu\lambda_{n-1} + \lambda_n = 0, \\ \qquad \lambda_{n-1} - 2\mu\lambda_n = 0. \end{cases} \tag{27}$$

The system (27) gives the recursive equations for the Chebyshev polynomials of the first kind T_i. Thus, we have:

$$\lambda_1 = \mu\lambda_0, \;\; \lambda_2 = (2\mu^2 - 1)\lambda_0, \;\; \ldots, \;\; \lambda_i = T_i(\mu)\lambda_0, \;\; i = 0, 1, \ldots, n.$$

The account of a norming condition gives the statement of Theorem 6.

LITERATURE CITED

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

2. Chapman D.G., Robbins H.E. Minimum variance estimation without regularity assumptions. — Ann. Math. Statist., **22**, 581-586, (1951).

3. Vil'chevskiy N.O., Shevlyakov G.L. Robust minimax estimation of a location parameter with a bounded variance. In: Stability Problems for Stochastic Models.Ed. by V.M. Zolotarev et al., Moscow / Utrecht, TVP / VSP, 279-288, (1994).