

# Selection Models Synthesis Based on Expert Estimates Extrapolation

Bugaev Y.V.\*

Voronezh State University of Engineering Technologies  
Voronezh, Russia  
e-mail: y\_bugaev52@mail.ru

Chikunov S.V.

Voronezh State University of Engineering Technologies  
Voronezh, Russia  
e-mail: chiksv@rambler.ru

Nikitin B.E.

Voronezh State University of Engineering Technologies  
Voronezh, Russia  
e-mail: nbe6419@gmail.com

Ivliev M.N.

Voronezh State University of Engineering Technologies  
Voronezh, Russia  
e-mail: max1m@mail.ru

**Abstract** — The analysis of large-scale business projects is non-dominant alternatives. However, the options under consideration may be too large, and the decision-maker may not be able to apply any mechanism for selecting the best option to this set. Most of the existing decision support procedures involve the entire available alternatives set in the comparison and evaluation process, so they are not suitable in this situation. The paper suggests an effective way to solve this problem – the expert assessments extrapolation method to develop an objective collective solution based on alternatives small training sample expert analysis. In the proposed method version, it is assumed that for any alternatives pair, experts are able to estimate the difference value in their utility. Thus, a difference-classification scale is introduced for alternatives, which makes it possible to more accurately assess the comparative alternatives value and make a more reasonable choice than when using an ordinal scale. This approach advantage also consists in the absence of any some alternatives superiority degrees priori numerical estimates over others, since any such assessment contains certain arbitrariness. The collective choice is based on obtaining generalized criterion parameters estimates using the maximum likelihood principle. In this case, calculating the likelihood function for m-alternatives sample requires determining the multiplicity  $m-1$  of several integrals numerical values over complex geometry region. It is proposed for its calculation to use the Monte-Carlo method. To increase the stability to the maximizing the likelihood function method integration error, we propose numerically-analytical method for calculating target function first and second orders partial derivatives. Simple and visible examples demonstrate the proposed approach effectiveness.

**Keywords** — *generalized criterion, maximum likelihood method, approximation, Monte-Carlo method.*

## I. INTRODUCTION

Developing and evaluating large business projects leads to the non-dominant alternatives set of problems. The options under consideration may be too large, and the decision-maker may not be able to apply any mechanism for selecting the best option to this set. The most well-known decision support procedures used all alternatives for comparing and evaluating. Therefore, in this case, known procedures cannot be used.

In order to solve this problem, the expert assessments extrapolation method (EAEM) can be applied [1, 2]. In this method, based on expert ordering alternatives from some small training sample, an equalities and inequalities system is constructed to describe the coefficient vector possible values range of the generalized criterion function. The method group version assumes that several experts are involved in evaluating alternatives. Their opinions discrepancies are considered as random errors and the generalized criterion coefficients final estimates are determined based on the maximum likelihood method. Here it is used the well – known Thurstone-Mosteller model assumptions [2], where expert estimates are interpreted as normally distributed random variables implementations. Currently, there are no other group selection procedures whose result was determined as the options small sample ranking result. Therefore, this approach seems to be a promising direction for constructing procedures for narrowing the set of non-worst options to acceptable sizes. Studies described in [1, 2] shows that the maximum likelihood method procedure is competitive in comparison with other collective choice rules (the rules of Bord, Copeland, Kemeny, and others).

In addition to comparing sample alternatives on the ordinal scale, it is proposed to use a stronger scale [1], which is called the difference-classification scale. The procedure assumes that for any alternative pair ( $A, B$ ), expert is able to estimate the value of its utility difference. Based on this assessment, the pair must be assigned to one of the classes  $Q_0, Q_1, \dots, Q_s$ . Each class is defined by a certain degree difference between  $A$  and  $B$ . Then, as in the usual expert assessments extrapolation method, equalities and inequalities system is constructed, which describes the generalized criterion function coefficients of vector possible values range. This expert assessments extrapolation method modification analysis [1] for small samples showed a significant improvement in the method characteristics compared to the examination option on the ordinal scale. However, in practice important problems involve the increase in the sample size, which is associated with the computational nature problems emergence. So, this work is devoted to solving the above problems.

II. MATHEMATICAL MODEL

The initial prerequisite for the expert assessments extrapolation method is the generalized criterion function existence:

$$F(x, b) = \sum_u b_u f_u(x) = b^T f(x), \tag{1}$$

where  $f_u(x)$  – known functions,  $b_u$  – unknown parameters (weights). According to the generalized criterion definition, an alternative is not worse than an alternative  $y (x \succeq y)$  if and only if  $F(x) \geq F(y)$ . Generalized criterion can be subjectively evaluated by an expert on an intuitive level. Function coefficients vector  $b$  possible values range (1) is determined from the training sample, based on expert alternatives comparison. Selection a single point from this area is the solution of this problem.

Expert assessments of extrapolation method can be used without generalized criterion usage. In this case, sample alternatives utility (value) estimates are determined directly. The ranking result of  $m$ -alternatives training sample by each  $r$ -expert can be represented as a matrix inequality

$$C(r) w \geq 0, \tag{2}$$

Where  $w$  – alternative values vector,  $C^{(r)}, r = 1, \dots, N$ – matrix displaying the ordering option selected by  $r$ -expert ( $r$ -order structural matrix),  $N$  – number of experts. The method for constructing matrices  $C^{(r)}$  is shown in the following example.

**Example 1.** There is four alternatives selection  $A_1, \dots, A_4$ . Assuming first expert ranked the sample on the ordinal scale as follows:  $A_2 \succeq A_1 \succeq A_3 \succeq A_4$ . Accordance this ordering we obtain an inequalities system for alternative values:

$$\begin{aligned} w_2 - w_1 &\geq 0; \\ w_1 - w_3 &\geq 0; \\ w_3 - w_4 &\geq 0. \end{aligned}$$

Thus, the matrix  $C^{(1)}$ , displaying the first expert ordering structure will have the form:

$$C^{(1)} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Assuming the second expert suggested the same rank on an ordinal scale. However, he further pointed out that the difference  $w_2 - w_1$  exceeds  $w_1 - w_3$ . The second expert used a stronger scale than the first expert, i.e. he used the difference-classification scale. This means that inequality is just:

$$(w_2 - w_1) - (w_1 - w_3) = -2 w_1 + w_2 + w_3 \geq 0.$$

Therefore, the matrix  $C^{(2)}$ , displaying the second expert ordering structure, will have the form:

$$C^{(2)} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ -2 & 1 & 1 & 0 \end{bmatrix}$$

The mentioned difference-classification scale is an intermediate variant between the ordinary ordinal scale and the hyper-ordinal scale. In this scale all the difference in the alternatives numerical evaluation are ordered. Using this scale these differences are divided into classes:

- alternatives pair  $(A, B)$  belong to the zero class  $Q_0$ - expert can't distinguish the elements of this pair by utility, i.e. the difference in their utility for him is close to zero;
- alternatives pair  $(A, B)$  belong to the first class  $Q_1$ - expert considers the alternative  $A$  to be more preferable than  $B$ , but this superiority is moderate;
- alternatives pair  $(A, B)$  belong to the second class  $Q_2$ - expert considers the alternative  $A$  more preferable than  $B$ , and this superiority is significantly stronger than in the first class pair, etc. So, in example 1, the second expert decided that  $(A_1, A_3) \in Q_1, (A_2, A_1) \in Q_2$ .

In our opinion, this approach advantage is the absence of any priority numerical estimates of some alternatives superiority degrees over others, since any such assessment contains certain arbitrariness. Hyper-ordinal scale usage does not seem reasonable to us, since it imposes excessive requirements on the expert resolution.

As it was mentioned above, the maximum likelihood method usage for obtaining numerical utility estimates is based on the expert evaluation results distribution independence and normality assumption. Then the  $i$ -alternative utility expert evaluation value is interpreted as a random variable with a mathematical expectation  $w_i$ , for which the desired statistical estimates will be obtained based on the maximum likelihood method results. We will also consider the sample to be homogeneous, which is why the estimated utility values have the same variance  $\sigma^2$ , which can be interpreted as the difference measure in expert opinions. Each proposed ranking option (2) is interpreted as a random event, the which probability is calculated using the formula

$$P_r(\theta) = \int_{D_r} g(x, w, \sigma) dx \tag{3}$$

$$D_r = \left\{ x \in E^m \mid C^{(r)} x \geq 0 \right\}$$

where  $g(x, w, \sigma)$  –  $m$ -dimensional normal distribution density form

$$g(x, w, \sigma) = \frac{1}{(2\pi)^{m/2} \sigma^m} \exp \left[ -\frac{1}{2\sigma^2} (x - w)^T (x - w) \right] \tag{4}$$

The likelihood function is built on these data

$$L(k_1, k_2, \dots, k_s, \theta) = \frac{N!}{k_1! \dots k_s!} P_1^{k_1}(\theta) \dots P_s^{k_s}(\theta), \tag{5}$$

where  $k_t$  – experts number, who selected the  $t$ -ordering variant,  $\sum_t k_t = N$ ;  $\theta = (w_1, \dots, w_m, \sigma)^T$  – estimated parameters vector.

Since the parameters  $w_i$  are in the indicator numerator, and  $\sigma$  –

is in the denominator, it's not the absolute values of these parameters that are important, but their ratio. Therefore,  $\sigma$  can be set to 1 and only search for the ratings values for  $w_i$ .

We can find the coefficients of maximum likelihood method estimates of the generalized criterion of function ranking alternatives adequately to the examination in a similar way. Let the vector  $x^i$  correspond to the  $i$ -alternative vector criterion evaluation and the function (1) known up to coefficients. Imposing restrictions  $F(x^i, b) = w_i$  and defining the likelihood function conditional maximum corresponding point, one can obtain the  $b_u$  coefficients maximum likelihood method estimates.

The success of the solution of this problem primarily depends on the procedure for calculating integrals (3) of successful choice, where the integration  $D_r$  area is generally a polyhedral cone. Using difference-classification scale, we will have a very complex integration area, and even with  $m \geq 4$  we will have to select special methods for solving the problem, individual for each examination result.

### III. RESULTS AND DISCUSSION

The use of the statistical test method (Monte-Carlo) convenience is that the numerical integration error order does not depend on the calculated integral multiplicity [3], but only on the random point's number. In addition, this method can be used when integrating across areas of almost any complex structure. However, when the likelihood function is maximized, the integrals (3) calculation is only an auxiliary task. Random component presence in the procedure for calculating integrals will inevitably lead to the fact that the function value will be calculated with a random error, and this fact will negatively affect the solution accuracy and the convergence rate.

In addition, the main problem is that the target function in this case will be discontinuous in optimization parameters terms. This is because when calculating integrals using the Monte-Carlo method, the integral function is averaged over the points that fall into the integration area, and the point's number that fall into the area changes abruptly during the desired maximization parameters variation.

However, we can offer a method that will work with a discontinuous target function. To be able to use this method, we need to define the target function first and second partial derivatives in analytical way

Let the generalized criteria function have the form (1), i.e. it depends linearly on the required coefficients. In this case, the ranking probability (3) is calculated using the formula

$$P(b) = \lambda \int_{C_{x \geq 0}} \exp\left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] dx$$

where  $\lambda$  – constant defined by the condition for normalizing the distribution density;  $Z$  – alternatives coordinates matrix in the basic functions space  $f_j$ , i. e.  $Z_{ij} = f_j(x^i)$ . If generalized criteria is not used, and only alternatives utility estimates are searched, then  $Z = I$ . As often happens when using maximum likelihood method, it is more convenient to search for the

likelihood function logarithm maximum. Therefore, we need to calculate a vector logarithmic derivative  $\frac{d \ln(P)}{db}$ , which weighted sum gives the likelihood function logarithm derivative. Then  $\frac{d \ln(P)}{db} = \frac{1}{P} \nabla_b P$ . In turn

$$\begin{aligned} \nabla_b P &= \lambda \int_{C_{x \geq 0}} \nabla_b \exp\left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] dx = \\ &= \lambda \int_{C_{x \geq 0}} \exp\left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] \times \\ &\quad \nabla_b \left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] dx \quad (6) \end{aligned}$$

The expression in the second square brackets can be converted to the form

$$\begin{aligned} -\frac{1}{2}(x - Zb)^T(x - Zb) &= \\ &= -\frac{1}{2}(x^T x - x^T Zb - b^T Z^T x + b^T Z^T Zb) = \\ &= \frac{1}{2}(-x^T x + (Z^T x)^T b + b^T Z^T x - b^T Z^T Zb). \end{aligned}$$

Matrix differentiation formulas are known:

$$\nabla_y (a^T y) = \nabla_y (y^T a) = a; \quad \nabla_y (y^T A y) = 2A y,$$

where  $a$  – constant vector,  $A$  – square symmetric matrix. From here

$$\begin{aligned} \nabla_b \left( -\frac{1}{2} \left[ -x^T x + (Z^T x)^T b + b^T Z^T x - b^T Z^T Zb \right] \right) &= \\ &= Z^T x - Z^T Zb = Z^T (x - Zb). \end{aligned}$$

Finally

$$\begin{aligned} \frac{d \ln(P)}{db} &= \frac{\lambda}{P} Z^T \times \\ &\int_{C_{x \geq 0}} \exp\left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] \cdot (x - Zb) dx \quad (7) \end{aligned}$$

Let us find out the probability-theoretic meaning of the formula (7). Consider the function

$$h(x) = \frac{1}{P} \lambda \exp\left[-\frac{1}{2}(x - Zb)^T(x - Zb)\right] =$$

$$= \frac{\exp\left[-\frac{1}{2}(x-Zb)^T(x-Zb)\right]}{\int_{Cx \geq 0} \exp\left[-\frac{1}{2}(x-Zb)^T(x-Zb)\right] dx}$$

It can be interpreted as the truncated multidimensional normal distribution with parameters  $(Zb, I)$  density defined in the region  $Cx \geq 0$ . Then the integral

$$\frac{\lambda}{P} \int_{Cx \geq 0} \exp\left[-\frac{1}{2}(x-Zb)^T(x-Zb)\right] \cdot (x-Zb) dx$$

represents the average value on the area  $C(v+Zb) \geq 0$  of the  $m$ -dimensional truncated normal random variable by the parameters  $(0, I)$  defined in this area. Let us denote  $\bar{v}$  – vector of this average value. Then the integral (6) is nothing but a vector  $Z^T \bar{v}$ . Its approximate value can be easily calculated using the Monte-Carlo method with the following algorithm.

1. Get  $K$  independent  $m$ -dimensional vectors  $v^i$  distributed according to the normal law with parameters  $(0, I)$ .
2. Select from them a set of  $D$  vectors, which satisfy the condition  $C(v^i + Zb) \geq 0$ .
3. Calculate the vectors  $v^i$  average value vector  $\bar{v}$ , get in  $D$ .
4. Calculate  $Z^T \bar{v}$ . This will be the value  $\frac{d \ln(P)}{db}$ , calculated using the statistical method.

To calculate the target function value, it is sufficient in this algorithm to calculate the vectors proportion in the  $D$ -region after point 2.

In the same way, we can calculate the second partial derivatives matrix and use it for optimization. For the record compactness, denote the integrand function as follows:

$$q(x, b) = \lambda \exp\left[-\frac{1}{2}(x-Zb)^T(x-Zb)\right]$$

Then

$$\begin{aligned} \frac{d^2 \ln(P)}{db^2} &= \frac{d}{db} \left( \frac{1}{P} \nabla_b P \right) = \\ &= -\frac{1}{P^2} \nabla_b P (\nabla_b P)^T + \frac{1}{P} \nabla_b (\nabla_b P) \end{aligned} \quad (8)$$

where  $\nabla_b P$  is calculated using (7).

We will use the following well-known formulas for matrix differentiation:

$$\nabla_y [f(y) \cdot Ay] = A \cdot [y \cdot (\nabla_y f)^T + f \cdot I] \quad (9)$$

where  $f(y)$  – vector argument  $y$  scalar function,  $A$  – constant matrix.

$$\nabla_y \left[ \vec{f}(y) \right] = \left[ \frac{\partial f_i}{\partial y_j} \right] = \left( \nabla_y \cdot \left( \vec{f} \right)^T \right)^T = \vec{f} \cdot (\nabla_y)^T \quad (10)$$

– Differentiation formula for a vector function  $\vec{f}(y)$ .

The latter formula is obtained by analogy with the finite increment formula for the vector function:

$$\vec{f}(x+h) \approx \vec{f}(x) + \nabla_x \left( \vec{f}(x) \right) \cdot h$$

Calculate  $\nabla_b (\nabla_b P)$ :

$$\begin{aligned} \nabla_b [\nabla_b P] &= \int_{Cx \geq 0} \nabla_b [q(x, b) Z^T x - q(b) Z^T Z b] dx = \\ &= \int_{Cx \geq 0} \nabla_b (q(x, b) Z^T x) - \nabla_b (q(x, b) Z^T Z b) dx \end{aligned}$$

Let us denote the last expression as  $\nabla_b$  and apply (9), (10) to it:

$$\begin{aligned} \nabla_b U &= \\ Z^T x (\nabla_b q(x, b))^T - Z^T Z b \nabla_b (q(x, b))^T - q(x, b) Z^T Z &= \\ = Z^T (x - Zb) (\nabla_b q)^T - q Z^T Z \end{aligned}$$

From here:

$$\nabla_b [\nabla_b P] = Z^T \int_{Cx \geq 0} [(x - Zb) (\nabla_b q)^T - q Z^T Z] dx \quad (11)$$

Finally, we get what  $\frac{d^2 \ln(P)}{db^2}$  is calculated by (8), where  $\nabla_b P$  is calculated by (7), and  $\nabla_b (\nabla_b P)$  – by (11). The integrals values are calculated using the Monte-Carlo method using pseudo-random numbers or the Sobol sequence [4].

The obtained formulas can be used for likelihood function approximate maximization by first-and second-order methods.

We will show the method application in the following simple example.

**Example 2.** Let three experts evaluate a two alternatives sample. Two experts were in ordering favor  $A_1 \geq A_2$ , and the third was in ordering favor  $A_2 \geq A_1$ . Replacing the variables, the corresponding likelihood function can be reduced to the form:

$$\ln L(a) = 2 \ln \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}t^2} \right) + \ln \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-\frac{1}{2}t^2} \right),$$

where  $a$  – desired parameter equal to the difference  $a = w_1 - w_2$ .

As a approximating integrals result using Sobol numbers at 127 points, the step function  $\ln L(a)$  (Fig. 1) was obtained.

To find its maximum, we use the Newton method at the initial approximation  $a_0 = 1.9$  with the derivatives calculation using (7), (8) and (11), where

$$P_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}t^2}, P_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-a} e^{-\frac{1}{2}t^2}, Z = 1.$$

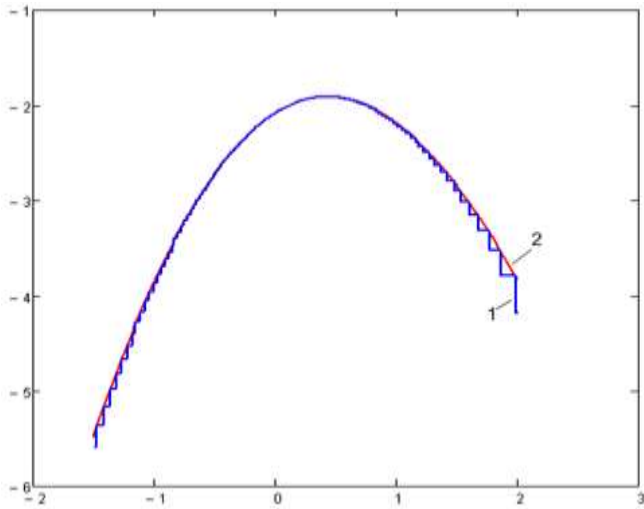


Fig. 1. 1 – approximation by the Sobol method; 2 – original function

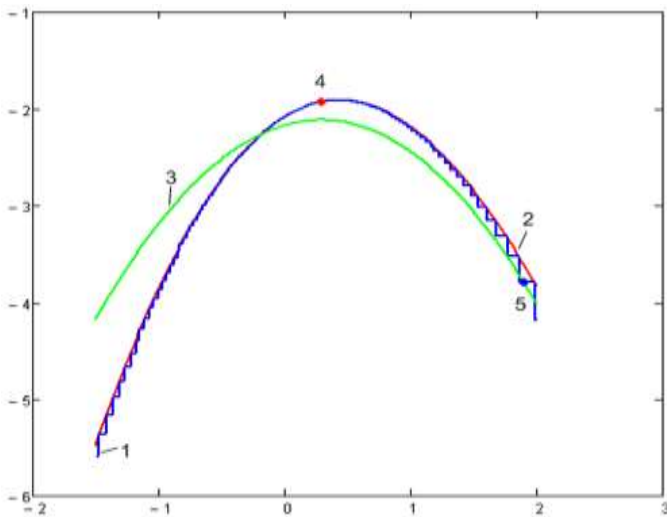


Fig. 2. 1 – target function approximation by the Sobol method; 2 – initial target function; 3 – target function parabolic approximation; 4 – first approximation to the optimal point; 5 – zero approximation.

On the 1<sup>st</sup> iteration, we get  $a_1 = 0.382$  (Fig. 2), and the result of the 2<sup>nd</sup> iteration  $a_2 = 0.436$ , as can be seen from Fig. 3, already almost coincides with the function maximum (the exact solution  $a^* = 0.431$ ).

Thus, the authors developed a multidimensional maximization program, based on the above material. This program implements the Newton method with step adjustment. This method was chosen because it is not sensitive to computational errors that occur due to the target function discontinuity. This is a positive difference from other methods – the conjugate gradients method, or variable metrics, where errors occur due to the need to calculate similar values

differences. In addition, in these methods the errors in them have a tendency to accumulate. In this version of the maximum likelihood method, the target function is strongly convex, and it is for optimizing such functions that the Newton method is well adapted. Group selection various applied tasks program approximation showed proposed approach high efficiency.

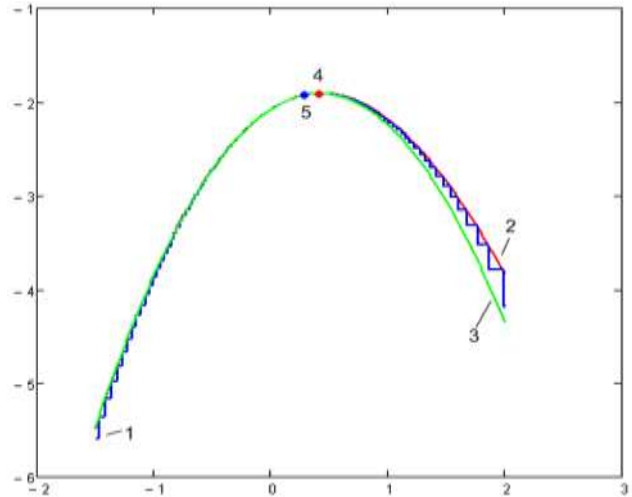


Fig. 3. 1 – target function approximation by the Sobol method; 2 – initial target function; 3 – target function parabolic approximation; 4 – first approximation to the optimal point; 5 – zero approximation.

**Example 3.** Let us consider the method possibilities on a more complex problem. Let there be  $m = 6$  alternatives sample, which was evaluated by 9 experts on the difference-classification scale. 6 of them ranked the sample as follows:  $(A_1, A_6), (A_2, A_4) \in Q_0; (A_5, A_1), (A_3, A_2), (A_5, A_6) \in Q_1; (A_4, A_5) \in Q_2$ .

Other three experts ranked the sample as follows:  $(A_1, A_6) \in Q_0; (A_5, A_1), (A_3, A_4), (A_5, A_6), (A_4, A_2), (A_2, A_5) \in Q_1$ .

Using maximum likelihood method result, sampling alternatives utility estimates were obtained. Since utility estimates are relative values for which only the relations between their differences are informative, it is convenient to bring them to the scale  $[0; 1]$ . The result is the following numerical values:  $w_3 = 1; w_4 = 0.839; w_2 = 0.686; w_5 = 0.313; w_6 = 0.0006; w_1 = 0$ . The likelihood function logarithm value at the optimal point was  $-0.00687$ , which corresponded to the resulting ordering probability 0.993. Such high probability is due to the fact that experts are unanimous in their assessments and disagreements are present only in details in general.

#### IV. CONCLUSION

In the course of the work, a method for finding point maximum likelihood method estimates was developed. These estimates were the Pareto-non-dominant alternatives generalized by criterion function coefficients, which were obtained from the small training sample expert ranking results. This method can be used for any expert ranking strategy, both on an ordinal scale and on a stronger difference – classification scale.

### **References**

- [1] U.V. Bugaev, M.S. Mironova, B.E. Nikitin, “Veroyatnostnii metod analiza procedur postroeniya kolektivnih ekspertnih ocenok”, Vest. VGU. Ser. Sistemnii analiz I informacionnie tehnologii, no. 2, pp. 130–135, 2011.
- [2] U.V. Bugaev, B.E. Nikitin, I.U. Shurupova, “Analiz veroyatnostnih svoistyv procedur postroeniya gruppovih ekspertnih ocenok”, Iskusstvennii intellekt I prinyatyie reshenii, no. 2, pp. 84–94, 2018.
- [3] S.M. Ermakov, Metod Monte-Carlo v vichislitelnoi matematike. St. Petersbourgh, 2009, 192 p.
- [4] I.M. Sobol, R.B. Statnikov, Vibor optimalnih parametrov v zadachah so mnogimi criteriyami. Drofa, 2006, 175 p.