Synergetic Aspects of Mathematical Statistics

ALBERT VORONIN[a], ALINA SAVCHENKO[b]

State University "Kyiv Aviation Institute", Kyiv
UKRAINE

[a]https://orcid.org/0000-0001-7201-1877,
[b]https://orcid.org/0000-0001-8205-8852

Abstract: - A comparative analysis of aspects characteristic of both mathematical statistics methods and synergetic data integration methods is carried out. The results of the analysis are used to improve the efficiency of statistical estimates calculated from a small sample, as well as to develop estimates of the characteristics of objects and processes in synergetic data integration systems with a limited number of channels.

Key-Words: - synergetic, data integration, degrees of freedom, discriminator, mathematical statistics, small sample.

1 Introduction

An important area of research in the field of synergetics [1,2] is the development of methods for data integration (merging). In complex synergetic systems, information about the same process (event) can be transmitted via several channels. The problem is to determine which channels transmit more reliable data and which transmit less reliable data. Then, two approaches can be taken. The first involves identifying one or more of the most informative (dominant) channels and cutting off the rest. This is done through the mechanism of "degrees of freedom reducers" [1]. This approach has the advantage of simplicity, but some useful nuances contained in the information of the cut-off channels are excluded from consideration and do not participate in the process of developing a cooperative solution.

The second way, in which useful information is not lost, seems more attractive. It is advisable to abandon the concept of the dominant when synthesizing a synergetic data integration system and, instead of "degrees of freedom reducers," include mechanisms that allow all data acquisition channels to participate in the decision-making process with weights corresponding to the degree of their informativeness in the current situation ("degrees of freedom discriminators"). As a result, all available information will be used appropriately.

The synergetic principle of data integration has much in common with the ideas of mathematical statistics [3]. Thus, the synergetic concept of data integration (merging) is used for the most reliable assessment of the characteristics of processes (objects) based on the available set of data. And mathematical statistics studies the methods of the most reliable assessment of the moments of distribution of random variables based on the available set of sample elements. The commonality of the problems of both theories makes the task of studying the synergetic aspects of mathematical statistics relevant both for synergetics and for the development of statistical methods.

The synergetic concept of data integration is actively used to extract maximum information from the available set of various data characterizing a process or an object in a wide variety of subject areas. The paper [4] describes a method for automatic classification of forest conditions based on aerospace survey materials based on the synergetic principle of data fusion. If the primary classification results in several different decisions regarding the class affiliation of an object, the overall decision is formed using a synergetic rule. First, the entire set of obtained decisions (a component of the synergetic system) is examined, after which the most informative (dominant) spectral channels of the sensor are determined, and the desired decision is made based on their readings. The paper [5] sets the task of combining signals from navigation fields of various physical natures (radio navigation fields such as GPS, geophysical fields, the field of stars and bodies of the Solar System, etc.) for the most reliable assessment of the current coordinates of the

spacecraft. The problem is solved by multi-criterial assessment of the suitability of each of the available fields using a nonlinear compromise scheme, after which the most informative (dominant) field in the current situation is selected. In the examples considered, the first way of data integration is used using the mechanism of "degrees of freedom reducers".

The second way is presented by the following works. In the monograph [6], the task of combining the readings of devices with different accuracy classes is set. In this case, each of the devices makes its contribution to the resulting reading in accordance with its accuracy class. Here the task of combining the data of expert assessments is solved taking into account the degree of competence of experts in the issue under consideration.

The method of expert assessment consists in the fact that to assess a certain quantitative characteristic, the postulates of not one, but several persons (experts) competent in the given issue are used. It is assumed that the "true" value of the quantitative characteristic unknown to us is within the range of expert assessments and the "generalized" collective opinion is more reliable. The unknown quantitative characteristic is considered as a random variable, the reflection of the distribution law of which is the expert's postulate. To establish the final assessment, the statements of all experts are studied in aggregate and processed as a kind of initial statistical material. The processing is carried out using the concepts of mathematical statistics.

The paper [7] describes a method of signal integration for bistatic radiolocation of small celestial bodies. To improve the accuracy of measurements when studying the motion parameters of small celestial bodies, a bistatic configuration of radar systems is used. Information from each of the receiving antennas, spaced at significant distances, is processed and compared with each other so that the resulting signal is the most reliable. In the examples given, the mechanism of "degrees of freedom discriminators" is used for data integration, and the data from the information transmission channels is considered as the initial statistical material.

Of particular interest is the case when the number of channels for obtaining the integrated data is limited, which corresponds to the case of a small sample in problems of mathematical statistics [8,9]. Since considerable experience has been accumulated in solving such problems, we will conduct further research in terms of mathematical statistics.

Estimates of the parameters of the probability distribution of a random variable are obtained based on the processing of statistical material, which is a set of experimental values of the random variable under study. The greater the amount of experimental data, the closer the population under consideration is to the general population (including all possible realizations of the random variable) and the more accurately the distribution parameters can be determined.

When solving problems of data integration, the researcher always has only limited statistical material at his disposal (a sample from the general population), and it is necessary to estimate the distribution parameters with the greatest possible accuracy. This is explained by the fact that obtaining each new element of the sample is usually a complex process associated with significant technical or economic difficulties. Computational difficulties play a less significant role. In such cases, O.K. Antonov said [10] that saving on calculations that evaluate huge economic events is the same as saving on aiming when shooting.

Therefore, the problem arises: to the maximum extent use information about the statistical properties of the random variable under study and obtain calculation algorithms for calculating refined estimates of distribution parameters based on statistical material of limited volume. Since the results of this study can be applied not only to increase the information content of data acquisition channels in complex synergetic systems, but also in other cases, it is advisable to formulate and solve the problem in general terms of mathematical statistics.

2 Problem Formulation

Let us consider a continuous real random variable X whose probability distribution density $f(x|\theta)$ is known up to an unknown parameter θ. A set of n independent realizations of the random variable X is given:

$$\bar{x} = \bar{x}^{(n)} = (x_1, x_2, ..., x_n). \quad (1)$$

(This is how the set of data to be integrated, obtained from n channels, is interpreted).

The problem is set: based on the results of a random sample (1), determine the best estimate θ^* in a certain sense of the unknown parameter θ of the distribution of a random variable X.

For example, if a random variable X is normally distributed with a known dispersion σ^2, then

$$f(x|\theta) = f(x|m_X) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(m_X - x)^2}{2\sigma^2}] \quad (2)$$

and the parameter to be estimated is the mathematical expectation: $\theta = m_X$.

As is known [11], the quality of statistical estimates is characterized by the following main properties:

1) consistency (convergence in probability of the estimate to the true value of the parameter);

2) unbiasedness (absence of systematic error in the estimate);

3) efficiency (minimum dispersion of the estimate).

If the researcher has only the information contained in the above statement of the problem, then the maximum likelihood method proposed by R. Fisher [3] is used to determine the best (consistent, unbiased and effective) estimates. Then the estimate of the mathematical expectation of a normally distributed random variable has the form

$$\theta^* = X_C = \frac{1}{n}\sum_{i=1}^{n} x_i. \quad (3)$$

Often, for small sample sizes, maximum likelihood estimates of type (3) do not provide satisfactory accuracy [8]. This forces the development of more effective estimation procedures in specific applications [3,8,12,13]. All of them, in one way or another, involve the use of additional information about the statistical properties of the random variable being studied.

3 Problem Solution

One of the most effective means of increasing the efficiency of statistical evaluation is the Bayesian approach [3]. It consists of the fact that the parameter θ to be estimated is considered as the realized value of the random variable Θ. The researcher expresses all the preliminary information available to him (before the experiments) in the form of an a priori distribution of the variable Θ, characterized by the probability distribution density $f_a(\theta)$. This function is called the a priori density and is considered known before the analysis of the data obtained experimentally. Bayes' theorem combines the a priori distribution and observational data to form an a posteriori distribution $f(\theta|x)$. In terms of the probability distribution densities for the random variables Θ and X, Bayes' theorem takes the form [3]:

$$f(\theta|x) = \frac{f(x|\theta)\, f_a(\theta)}{f(x)}, \quad (4)$$

where the marginal distribution $f(x)$ is expressed by the formula

$$f(x) = \int_{-\infty}^{+\infty} f(x|\theta)\, f_a(\theta) d\theta.$$

The physical meaning of Bayes' theorem in the form (4) is that if $f_a(\theta)$ is the probability density function assigned to the parameter θ before conducting the experiments, then $f(\theta|x)$ is the density function that should be assigned to θ after obtaining the data. In terms of the synergetic theory of data integration, this means using the mechanism of "degrees of freedom discriminators". Statistical estimates calculated on the basis of the posterior distribution are of better quality than maximum likelihood estimates, since they use additional information about the unknown parameter θ in the form of the prior distribution $f_a(\theta)$.

The most delicate point in Bayesian parameter estimation is the assignment of the prior density function. It must be adequate to the available preliminary information. On the one hand, in no case should one introduce information that is not present

in the prior data. Therefore, one selects $f_a(\theta)$ based on the requirement that it have the maximum possible entropy (in the Shannon sense [14]) under the given conditions in the form of specific prior data considered as constraints [8]. On the other hand, neglecting any objective prior information leads to the selection of a less informative prior density, which makes the statistical estimate less effective.

For example, let us consider the case of normal distribution of a random variable *X* with an unknown mathematical expectation m_x and a known variance σ^2, as presented by formula (2). It is known [8] that an unbiased estimate of the mathematical expectation, considered as a random variable, is also distributed according to a normal law with variance σ^2/n and with the same mathematical expectation m_x. Since the parameter m_x is subject to estimation and is unknown, then the usual Bayesian practice prohibits choosing a normal distribution law as $f_a(\theta) = f_a(m_x)$. Thus, objective a priori information about the type of distribution law of the estimated parameter is lost, which negatively affects the effectiveness of the statistical estimate.

It is proposed in such cases to introduce an a priori density with the parameter θ', considered as an unknown constant: $f_a(\theta|\theta')$. For our example with a normal distribution

$$f(m_x|\theta') = \frac{1}{(\sigma^2/n)\sqrt{2\pi}} \exp\left[-\frac{(\theta' - m_x)^2}{2(\sigma^2/n)}\right].$$

Such a priori density does not introduce information that is not present in the a priori data, and at the same time allows the use of objective a priori information about the type of distribution law (in our case, normal) of the estimated parameter.

Let us find an algorithm for calculating a refined estimate of the parameter θ. For each random sample $\bar{x}$, the sought computational algorithm $\theta(\bar{x})$ will give an estimate θ^* of the unknown parameter θ. If we use the value θ^* while the true value is θ, then an error occurs, the cost of which can be expressed as a loss function $c(\theta^*, \theta)$. The choice of the loss function is not determined by theory, is subjective in nature and expresses the attitude of the researcher to the magnitude of the discrepancy between the decision θ^* made and the true value of the estimated parameter θ [13]. The loss function must be non-decreasing.

Since the value of θ is unknown, it is impossible to calculate the true loss function. However, using a priori information in the form of the a priori density $f_a(\theta|\theta')$ we proposed, for all possible realizations of the observation vector $\bar{x}$ with known statistics $f(x|\theta)$, we can introduce the risk function

$$R(\theta^*) \equiv M[c(\theta^*, \theta)] = \int_{-\infty}^{+\infty}\int c(\theta^*, \theta) f(x|\theta) f_a(\theta|\theta') dx d\theta, \quad (5)$$

which is defined as the mathematical expectation of the loss function. Using Bayes' theorem in form (4), we write expression (5) as

$$R(\theta^*) = \int_{-\infty}^{+\infty}\int c(\theta^*, \theta) f(\theta|x) f(x) dx d\theta.$$

We obtain the optimal estimate θ^* by minimizing the risk function:

$$\theta^* = \arg\min_{\theta^*} R(\theta^*),$$

where

$$\min_{\theta^*} R(\theta^*) = \min_{\theta^*} \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} c(\theta^*, \theta) f(\theta|x) d\theta] f(x) dx.$$

For a fixed sample $\bar{x} = \bar{x}^{(n)}$, only the expression in square brackets, called the conditional Bayesian risk $r(\theta^*)$, depends on the parameter θ^* [12]. Therefore, minimizing the risk function $R(\theta^*)$ is completely equivalent to minimizing the conditional Bayesian risk $r(\theta^*)$ for a given sample $\bar{x} = \bar{x}^{(n)}$:

$$\min_{\theta^*} R(\theta^*) = \min_{\theta^*} r(\theta^*)\Big|_{\bar{x}=\bar{x}^{(n)}} = \\ = \min_{\theta^*} \int_{-\infty}^{+\infty} c(\theta^*,\theta) f\left(\theta\middle|x\right) d\theta\Big|_{\bar{x}=\bar{x}^{(n)}}. \quad (6)$$

Further derivation requires specification of the loss function. In estimation practice, the quadratic function is usually used

$$c(\theta^*,\theta) = (\theta - \theta^*)^2. \quad (7)$$

Using the necessary condition for the minimum of the function

$$\partial r(\theta^*)/\partial\theta^* = 0$$

we express (6) taking into account (7) as

$$2\int_{-\infty}^{+\infty} (\theta^* - \theta) f\left(\theta\middle|x\right) d\theta\Big|_{\bar{x}=\bar{x}^{(n)}} = 0.$$

Let's transform this expression into the form

$$\theta^* \int_{-\infty}^{+\infty} f(\theta|x) d\theta = \int_{-\infty}^{+\infty} \theta f(\theta|x) d\theta\Big|_{\bar{x}=\bar{x}^{(n)}}.$$

Considering that according to the property of probability density

$$\int_{-\infty}^{+\infty} f(\theta|x) d\theta = 1,$$

we write it down finally

$$\theta^* = \int_{-\infty}^{+\infty} \theta f(\theta|x) d\theta\Big|_{\bar{x}=\bar{x}^{(n)}}, \quad (8)$$

i.e., with a quadratic loss function, the optimal estimate θ^* is the a posteriori mathematical expectation of the parameter θ, calculated based on a given vector of observations.

We will use the definition of the posterior density in the form (4) and transform expression (8) to the form

$$\theta^* = \frac{\int_{-\infty}^{+\infty} \theta f(\theta|x) f_a(\theta\middle|\theta') d\theta}{\int_{-\infty}^{+\infty} f\left(\theta\middle|x\right) f_a(\theta\middle|\theta') d\theta}\Big|_{\bar{x}=\bar{x}(n)}. \quad (9)$$

Since the desired estimate $\theta^* = \theta^*(x_1, x_2, ..., x_n)$ must be calculated based on a given vector of observations, we must move in expression (9) from integrals to summation over the elements of a given sample and replace the unknown constants with their estimates:

$$\theta^* = \frac{\sum_{i=1}^{n} x_i f(x_i\middle|\theta^*) f_a(x_i\middle|\theta^*)}{\sum_{i=1}^{n} f(x_i\middle|\theta^*) f_a(x_i\middle|\theta^*)}. \quad (10)$$

Formula (10) expresses the dependence

$$\theta^* = \varphi(x_1, x_2, ..., x_n; \theta^*).$$

As is known [15], an equation in this form can be solved by an iterative method. The iterative procedure is organized in accordance with the recurrent formula

$$\theta^*[l] = \varphi(x_1, x_2, ..., x_n; \theta^*[l-1]), l \in [1, L],$$

and the iterative process ends when the condition is met

$$\theta^*[L] - \theta^*[L-1] \le \lambda_\theta,$$

where l is the number of the current iteration; λ_θ is the specified accuracy of the estimate calculation θ^*. If it is necessary to analyze convergence issues, then one can apply the well-known theorem [15], according to which for the convergence of the iterative process it is sufficient that the inequality is satisfied on the considered interval of estimate θ^* refinement

$$\left|d\varphi(x_1, x_2, ..., x_n; \theta^*)/d\theta^*\right| < 1.$$

E x a m p l e . Let us consider the problem of refined statistical estimation of the mathematical expectation of a random variable X distributed according to the normal law with the distribution density $f(x|\theta) = f(x|m_x)$ given by formula (2), based on the results of a random sample (1), if the variance σ^2 is known. The a priori information is that the estimate of the mathematical expectation X_c is also distributed according to the normal law with a known variance

$$\sigma_1^2 = \sigma^2 / n \qquad (11)$$

and with the same unknown mathematical expectation m_x:

$$f_a(\theta|\theta') = f_a(m_x|X_c) = \\ = \frac{1}{\sigma_1\sqrt{2\pi}} \exp[-\frac{(X_c - m_x)^2}{2\sigma_1^2}]. \qquad (12)$$

To solve the problem, we substitute expressions (2) and (12) into formula (10) and obtain

$$X_C = \frac{\sum_{i=1}^{n} x_i \exp[-\frac{(x_i - X_C)^2(\sigma^2 + \sigma_1^2)}{2\sigma^2\sigma_1^2}]}{\sum_{i=1}^{n} \exp[-\frac{(x_i - X_C)^2(\sigma^2 + \sigma_1^2)}{2\sigma^2\sigma_1^2}]},$$

from which the effect of reducing the variance of the posterior distribution is visible, which indicates the effectiveness of the proposed estimation procedure. Taking into account (11), we obtain

$$\frac{(\sigma^2 + \sigma_1^2)}{2\sigma^2\sigma_1^2} = \frac{n+1}{2\sigma^2}.$$

Thus, an iterative procedure must be organized to calculate the estimate X_C.

$$X_C[l] = \frac{\sum_{i=1}^{n} x_i \exp[-\frac{(x_i - X_C[l-1])^2(n+1)}{2\sigma^2}]}{\sum_{i=1}^{n} \exp[-\frac{(x_i - X_C[l-1])^2(n+1)}{2\sigma^2}]},$$

(13)

at that, as a first approximation it is advisable to adopt the maximum likelihood estimate (3):

$$X_C[l] = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

For quantitative verification of the proposed iterative algorithms, tables of normally distributed random numbers [16] with known distribution parameters were used. The calculation results are given in [6].

The results show that the confidence interval corresponding to the refined estimate is smaller than the confidence interval of the maximum likelihood estimate. The greatest gain in efficiency is obtained with small sample sizes. This confirms the idea that with an increase in the volume of measurements, the relative contribution of a priori information in obtaining estimates becomes smaller and smaller, and the Bayesian estimate and the maximum likelihood estimate asymptotically coincide [17]. Therefore, it is advisable to calculate the refined estimate mainly with small sample sizes.

An important property of the a priori density is that it should not be its own density, i.e. its integral does not necessarily have to be equal to one [3]. In a number of cases [12], attempts to use pseudo-Bayesian estimates are considered entirely justified, in the construction of which some other density is introduced instead of the missing a priori probability density of the estimated parameter. Of particular interest is the possibility of using any of the so-called potential functions as a priori density [18], a special case of which is the normal distribution law (2). Other examples include the functions

$$f_1 = \frac{\alpha}{|X_c - x|},\ f_2 = \frac{\beta}{(X_c - x)^2},\ f_3 = \frac{\gamma}{1 + \delta(X_c - x)^2},$$

(14)

where $\alpha, \beta, \gamma, \delta$ are constants. The potential function is characterized by the fact that it monotonically decreases with distance from the value X_C , i.e. it is symmetrically even with respect to X_C . If it is only known that the estimated parameter is distributed symmetrically in the general population, then it is advisable to obtain a more accurate estimate X_C by choosing a sufficiently simple potential function for the a priori density.

Sometimes, to reduce the amount of calculations, it is even convenient to deliberately replace a known (for example, normal) distribution law with another, simpler potential function. Thus, if a random variable is distributed according to a uniform law, then it is known that the estimate of its mean value obeys the normal distribution law. However, having chosen the first of the potential functions (14) as the a priori density, we arrive at the following simple iterative algorithm for calculating the estimate X_C :

$$X_C[l] = \sum_{i=1}^{n} \frac{x_i}{|X_C[l-1] - x_i| \sum_{j=1}^{n} \frac{1}{|X_C[l-1] - x_j|}},$$

$$X_C[1] = \frac{1}{n}\sum_{i=1}^{n} x_i; l \in [1, L], X_C[L] - X_C[L-1] \le \lambda_x. \quad (15)$$

Algorithm (15) in compact form represents a method for processing expert assessment data, described in detail in [6].

The general expression for the refined estimate (10) is fully consistent with Gauss's idea [12] that the most probable value of the estimated parameter is the one that minimizes the sum of the squares of the differences between the actually observed and calculated values, multiplied by the weighting coefficient k_i , reflecting the relative confidence in the observations:

$$\theta^* = \arg\min_{\theta^*} \sum_{i=1}^{n} k_i (x_i - \theta^*)^2. \quad (16)$$

In [19,20] it is shown that expression (10) is indeed obtained from (16) if the posterior probability density function ("degrees of freedom discriminator") is introduced as a measure of relative confidence in observations.

Thus, the proposed method provides for an individual approach to each realization of a random variable (weighting in accordance with the a posteriori probability of its occurrence), which allows [8] to eliminate information loss when calculating the desired estimates for a small sample.

It is important to note that the most reliable assessment is developed through the organization of an iterative process in which the elements of the sample interact with each other at each iteration. Similarly, synergetics provides for a process characterized by self-management and self-organization in accordance with the set goal. Here, complex processes develop through the collective interaction of components. Cooperation of components allows the use of reserve capabilities of the system and significantly increases the degree of emergence (system effect).

4 Conclusion

A comparative analysis of aspects characteristic of both mathematical statistics methods and synergetic data integration methods was conducted. The results of the analysis are used to improve the efficiency of statistical estimates calculated from a small sample, as well as to develop the most reliable estimates of the characteristics of objects and processes in synergetic data integration systems with a limited number of channels.

In the future, it is planned to study the possibility of using synergetic weighting in the least squares method to increase the convergence rate of the dual programming method described in [21]. It is planned to conduct a study of the possibility of using synergetic methods in various subject areas.

References:

[1]. Kolesnikov A.A. Synergetic Theory of Control. – Moscow: Energoatomizdat, 1994. – 344 p.
[2]. Haken G. Synergetics. – Moscow: Mir, 1980. – 248 p.
[3]. Cox D., Hinkley D. Theoretical Statistics. – Moscow: Mir, 1978. – 560 p.
[4]. Lyalko V.I., Fedorovsky A.D., Popov M.A. et al. Using Satellite Imagery Data to Study Natural Resource Problems // Space Research in Ukraine (2002-2004) – Kyiv: NKAU, 2004. – P.7-14.

[5]. Varlamov I.D., P'yaskovsky D.V., Vodop'yan S.V. Adaptive correlation-extremal algorithm for spacecraft navigation in geophysical fields based on differential-Taylor transformations // Space Science and Technology. - 2001. - No. 4. - P. 141-146.
[6]. Voronin A.N., Ziatdinov Yu.K., Kharchenko A.V. Complex technical and ergatic systems: research methods. - Kharkov: Fact, 1997. - 240 p.
[7]. Voronin A.N. Method of signal integration for bistatic radar of small celestial bodies // Abstracts of reports of the 9th international conference "System Analysis and Control". – M.: MAI Publishing House, 2004. – P.113-114.
[8]. Gaskarov D.V., Shapovalov V.I. Small Sample. – M.: Statistika, 1978. – 248 p.
[9]. Fedolov A.A., Fedolov Yu.G., Tsygichko V.N. Introduction to the Theory of Statistically Unreliable Decisions. – M.: Statistika, 1979. – 279 p.
[10]. Beshelev S.D., Gurvich F.G. Expert Assessments. – M.: Nauka, 1973. – 127 p.
[11]. Ventzel E.S. Probability Theory. – M.: Nauka, 1969. – 576 p.
[12]. Sage E., Mels J. Estimation Theory and Its Application to Communications and Control. Moscow: Svyaz, 1976. 496 p.
[13]. Mostovoy S. V. Optimal Estimates of Seismic Wave Field Parameters. Kiev: Nauk. Dumka, 1979. 112 p.
[14]. Shannon K. Works on Information Theory and Cybernetics. Moscow: Foreign Literature Publishing House, 1963. 829 p.
[15]. Guter R. S., Reznikovsky P. T. Programming and Computational Mathematics. Moscow: Nauka, 1971. Issue 2. 273 p.
[16]. Shannon R. Simulation Modeling of Systems: Art and Science. Moscow: Mir, 1978. 419 p.
[17]. Mudrov V.I., Kushko V.L. Methods of processing measurements. - M.: Sov. radio, 1976. - 192 p.
[18]. Aizerman M.A., Braverman E.M., Rozonoer L.I. The method of potential functions in the theory of machine learning. - M.: Nauka, 1970. - 384 p.
[19]. Voronin A.N. On the increase in the efficiency of statistical estimates for parameters of ergatic systems // Cybernetics and Computational Technology. - 1980. - Issue 50. - P. 29-31.
[20]. Voronin A.N. On the rise of efficiency of statistical estimates for parameters of ergatic systems // Zentralblatt fur Mathematik und ihre Grenzgebiete. Mathematics Abstracts. - Band 484. - Berlin. Heidelberg. New York. – 24.01.1983. – R. 375.
[21]. Voronin A.N. Adaptive approximation models in optimization problems // Cybernetics and systems analysis. – 1994. – No. 5. – P. 83-93.