ISSN 2707-1898 (print)
Український журнал інформаційних технологій
Ukrainian Journal of Information Technology
http://science.lpnu.ua/uk/ujit
https://doi.org/10.23939/ujit2019.01.058
Article received 27.10.2019 р.
Article accepted 20.11.2019 р.
UDK 004.413:519.[9+81]

Correspondence author
Yu. I. Hrytsiuk
yurii.i.hrytsiuk@lpnu.ua

Yu. I. Hrytsiuk, O. R. Ferneza
Lviv Polytechnic National University, Lviv, Ukraine

REFINING EXPERT BASED EVALUATION ON THE BASIS OF A LIMITED QUANTITY OF DATA

A technique has been developed to refine expert based evaluation of the probability distribution parameter of a random variable based on a limited amount of statistical data. This made it possible to identify the most informative data transmission channel (the most qualified expert) and get its reliable assessment. It has been established that the analysis and processing of a limited amount of data is carried out using well-known techniques in probability theory and mathematical statistics, where significant theoretical and practical experience has been accumulated. A mathematical model that describes the state of an object, process, or phenomenon is presented as a point estimate of the probability distribution parameter of a random variable, the value of which is obtained on the basis of a small sample of data. The modern approaches to the statistical estimation of a random variable are analyzed, the most common of which is the Bayesian approach. It is established that the most significant moment of the Bayesian estimation of the unknown parameter of the probability distribution of a random variable is the appointment of a certain function of the a priori density of its distribution. This function should correspond to the available preliminary information on the shape of the a priori probability distribution of this quantity.

The traditional approach to identifying the most informative channel for transmitting data on the state of an object, the course of a process or phenomenon, and cutting off others is less reliable. This is carried out using the so-called mechanism of reducers of degrees of freedom. Its main disadvantage is that in the cut-off data transmission channels, there may be some useful information that is not involved in the development of an agreed solution. Therefore, it is necessary to introduce mechanisms of discriminators of degrees of freedom. They allow all data transmission channels to participate in the decision-making process in terms of importance, which corresponds to the greatest degree of their information content in the current situation. An illustrative example of the application of the considered methods of averaging data is shown, which reflects the results of calculations by iterations using the implementation mechanisms of both reducers and discriminators of degrees of freedom. These mechanisms reflect the features of the implementation of iterative algorithms that are characteristic of both methods of mathematical statistics and methods of a synergetic system of averaging data.

***Keywords*:** probability theory; mathematical statistics; data averaging methods; informative data transmission channel; mechanism of reducers of degrees of freedom; the mechanism of discriminators of degrees of freedom; iterative algorithms.

Introduction

Software quality is one of its main characteristics in various areas of the application of information technology [22]. Evaluating software quality usually means the procedure that determines its properties, designation and effective operation. This approach to software evaluation becomes of particular importance with the development and improvement of technology of expert based data analysis [20]. Hence, there is the need for developing methods and tools for comprehensive evaluation of various indicators of software quality that would consider both uncertainties in the input data and subjectivity of expert based evaluation [4], [34].

Existing methods for determining software quality on the basis of subjective expert based evaluation are usually designed assuming that their distribution is normal [3], [5], as many random variables are subject to this law. The assumption of the normal distribution of expert based evaluation to identify software quality indicators often appears to be false [24], [31]. In practice, some researchers have to deal with selective parameters of some random value that are determined by a small number of values of the parameter studied [9], [12]. They are considered the point estimation of distribution parameters and they are random. Therefore, if the distribution of experimental data including expert based evaluation significantly differs from the normal law, it is necessary to use somewhat different methods and ratios to calculate the relevant indicators [8], [31], [32].

However, when solving many problems of mathematical statistics, researchers usually have only limited statistical data available (a small sample of the total amount), and the point estimation of θ parameter of probability distribution of the random variable X requires the greatest accuracy possible [18]. The reason is that getting every new sample item is usually a complex process associated with significant difficulties of technical or economic nature. At the same time, the difficulties associated with the implementation of the relevant calculations are much less complex and their implementation cost is low [26].

Analysis of recent research and publications. In [26] the authors substantiate the averaging method to stochastic systems with fast changing phase, to which oscillating processes in significantly nonlinear systems are reduced. The work also provides formulation of the theorem to evaluate the difference between the solutions of exact and averaging in the first approximation systems on a finite time interval. In [19] the programming model for calculating the average value of noise voltage square by the method of averaging by units. In [27] they offer a modification of the method of barycentric averaging, which allows formalizing the procedure of the selection of position of individual computer templates and limitation of their number.

However, the papers mentioned do not provide a comparative analysis of the peculiarities of the implementation of calculation algorithms specific both for the methods of mathematical statistics and the methods for synergistic system of data averaging. The advantages of some of them and disadvantages of the others are not given as well. Therefore, the study is relevant, which would allow getting calculation algorithms for specified point evaluation of the probability distribution parameter of a random value based on a limited amount of statistical material using available information about the statistical properties of the random value investigated.

The approach to data averaging proposed in [8] enables obtaining maximum amount of information available with a limited number of communication channels. The author uses the so-called discriminators of the freedom degrees along with the concept of reducers of freedom degrees that allow all communication channels to participate in producing of an agreed solution depending on the level of their information content in the current situation [32]. We will try to adapt these techniques for mathematical averaging of expert based evaluations of software quality, which allow identifying the most qualified expert and get his accurate evaluation.

Object of the study is data averaging with a limited number of channels.

Subject of the study is methods and tools for specifying point evaluation of the probability distribution parameter of a random value based on a limited amount of statistical data.

The purpose of the study is to develop the methods of specified point evaluation of the probability distribution parameter of a random value based on a limited amount of statistical data that will help identify the most informative data channel and get it accurate evaluation.

To accomplish this goal, you must complete the following basic tasks:

1) give examples that demonstrate the need to determine the relative degree of reliability of data received by different communication channels about the same state of the object, the course of the process or phenomenon and the need for their refinement;
2) analyze current approaches to the statistical estimation of a random variable, which will allow the assignment of a certain function of the a priori density of probability distribution of this value;
3) give an illustrative example of the application of data refinement mechanisms, which reflects the results of calculations by iterations using the so-called mechanisms of realization of reducers and discriminators of degrees of freedom;
4) make appropriate conclusions and make recommendations on the use of the developed methodology for refining the expert assessments of the software quality on the basis of a limited amount of statistics.

1. Examples of specifying of average values

In modern software systems data describing the state (x) of the same object, process or phenomenon can be transmitted through several channels of communication (n). The problem is to determine the relative degree of reliability of data received from each communication channel at the current period of time, and to develop a mechanism for averaging the most reliable evaluation of x^* actual state of the object x, the course of process or phenomenon with a limited number of communication channels [17]. Let us consider some examples that demonstrate the need to determine the relative degree of data reliability and the need of its averaging [8].

Example 1. Averaging of indicators of tools with different accuracy class (the problem is set in [16]). Herewith, each tool adds some disturbance in the resulting indicator according to its accuracy class.

Example 2. Averaging of signals for bistatic radiolocation of small celestial bodies (the method described in [8]). To improve the accuracy of measurements in the study of motion parameters of these bodies, bistatic configuration of radar systems is used. Data from each of the receiving antennas separated by considerable distances from one another is subjected to processing and comparing them with each other so that the resulting signal is the most reliable.

Example 3. Averaging experts based evaluation considering their degree of competence in this matter [34]. The method of expert based evaluation is to determine that some quantitative characteristics of the object, phenomenon or process use statements of several experts instead of one who are competent in the subject knowledge area. It is assumed that the true meaning of unknown quantitative characteristics is somewhere in the middle range of expert based evaluations, i.e. their generalized collective opinion is more reliable.

Example 4. True software quality assessment can be based on several criteria: 1) software focused on the end user; 2) software focused on process of execution and results obtained; 3) management of users' participation in the software performance and their liability for it; 4) continuous improvement of software functionality options; 5) software problems that depend on users; 6) measurement of software efficiency; 7) team work organization to improve software quality. These criteria are absolutely correct both for software as a product and its development as a process. Data on software quality according to each of these criteria is provided by several communication channels (from several experts), it is specified, i.e. the most reliable assessment of the true software quality is made.

In the examples provided for data averaging they consider unknown quantitative characteristic of the state x of the object, process or phenomenon as a random variable, where a signal from the communication channel represents the distributive law. In order to provide a final evaluation, data obtained from all channels are collectively analysed and processed as some initial statistical material. Analysis and data processing are usually carried out applying known techniques of probability theory and mathematical statistics [19], [20], [35], where considerable theoretical and practical experience is accumulated.

In everyday practice, the number of data communication channels for its further averaging is relatively small, similarly as in the case of small size sample in problems of mathematical statistics [22], [23], [34]. Thus, the model of the state of the object, process or phenomenon can be represented as the point evaluation of the probability distribution parameter of a random value, which is obtained based on the limited amount of the sample [18], [33].

Considering Example 4, interactively collected assessment of software quality provided by each expert is stored in database which can be accessed through the appropriate software tool. Expert evaluations should be conducted in the form of a survey using a ranged scale for each of the criteria. Consequently, each of the experts should make appropriate assessments, which are then inserted into the corresponding Table 1. Each of the experts will influence software quality for each criterion and a generalized indicator of its quality in different ways.

Table 1. Criteria for evaluation of software quality and related expert based opinions

Num ber	Criteria for evaluating software quality	Experts:					Average value
		1	2	3	4	5	
		Experts estimate, points					
1	Software focused on the end user	7.9	9.6	9.2	8.7	9.9	9.06
2	Software focused on process of execution and results obtained	8.3	7.8	9.1	9.8	9.2	8,84
3	Management of users' participation in the software performance and their liability for it	7.1	9.8	7.6	7.9	6.9	7.86
4	Continuous improvement of software functionality options	6.4	7.8	7.5	6.9	8.1	7.34
5	Software problems that depend on users	8.3	9.1	8.7	9.0	9.7	8,96
6	Measurement of software efficiency	7.4	9.8	8.6	9.2	8.4	8.68
7	Team work organization to improve software quality	8.7	9.5	9.6	7.6	9.4	8,96
	Total score	54.1	63.4	60.3	59.1	61.6	**59.70**

Since the results of this study can be applied not only to increase the in formativeness of potential data transmission channels for modern software systems, but also in other areas of scientific, industrial and economic activity, further formulation and solution of problems should be performed in general terms of mathematical statistics [22], [23].

The main task of mathematical statistics is to find a distribution of the investigated random variable X according to the current sampling [8]. In many cases, the form of distribution of variable X can be considered to be known and researcher's task is to determine the unknown parameter θ of this distribution. Moreover, the point evaluation of the parameter θ of probability distribution of the random variable X can be determined on the basis of statistical data processing, which consists of a set of experimental values of the investigated random variable [19], [20].

Thus, we consider actual continuous random variable X, which probability distribution density function $f(x|\theta)$ is known with accuracy up to an unknown parameter θ. A set of independent implementation of a random variable X is given in the following set:

$$\tilde{X} = \{x_j,\ j = \overline{1,n}\}\,, \tag{1}$$

where x_j is the value of the i-th random variable; n is the number of independent implementations. The sample (1), which interprets the data set to be the subject to averaging, is obtained n from transmission channels. According to this sampling we should define probably the best point evaluation $\hat{\theta}$ of the unknown parameter θ of probability distribution of the random variable X. If the random variable X is normally distributed with known variance σ^2, then according to the law of normally distributed random variable

$$f(x|\theta) = f(x|m_x) = \frac{1}{\sigma\sqrt{2\pi}} Exp\left(-\frac{(m_x - x)^2}{2\sigma^2}\right) \tag{2}$$

and parameter θ, to be assessed, is mathematical expectation: $\theta = m_x$.

According to the information given in [29], the quality of statistical estimates has the following main features: 1) *capacity*, if with increasing sample size n is likely to follow the estimated parameter; 2) *unshiftedness* if its expected mathematical value is the parameter being estimated; 3) *efficiency* when the dispersion is minimal for a given sample size.

If the researcher has only the information contained in the above formulation of the problem, then in order to determine the best (capable, effective, and unshifted) estimates of unknown parameter θ, maximum likelihood estimation is applied, which was proposed by the British statistician Ronald Fisher in 1912 in [32]. Then mathematical expectation estimate of normally distributed random variable is as follows:

$$\hat{\theta} = X_c = \frac{1}{n}\sum_{j=1}^{n} x_j\,, \tag{3}$$

where X_s is the average value of the sample (1), which reflects the mathematical expectation estimate m_x.

For small samples, maximum likelihood estimates obtained by the formula (3), do not often provide the researcher with satisfactory accuracy [35]. This requires developing of more effective evaluation procedures of the parameter θ in order to give reliable mathematical expectation m_x in specific software applications [2], [33]. All of them are partly related to the use of additional information on the statistical properties of the investigated random variable.

Thus, there are some examples that demonstrate the need to determine the relative degree of reliability of data obtained from various communication channels and the necessity of their averaging. We have revealed that data is analysed and processed using known techniques of probability theory and mathematical statistics, where considerable theoretical and practical experience is accumulated. As the result of data analysis and processing, the model of the state of the object, process or phenomenon can be represented as the point evaluation of the probability distribution parameter of a random value, which is obtained based on the limited amount of the sample.

2. Current approaches to statistical estimation of random variable

An effective means to enhance the quality of statistical evaluation of software quality by various criteria is Bayesian approach [2]. The physical meaning of Bayes theorem is that if the function $f_a(\theta)$ is a priori density probability distribution of the random variable that is considered to be a parameter θ before the procedure of software quality evaluation, the function $f(x|\theta)$ will be a posteriori probability of this value, which should attribute parameter θ after obtaining the data. Statistical expert based evaluation calculated based on a posteriori probability distribution, is of better quality than maximum likelihood estimation, as it uses additional information on the unknown parameter θ as a priori density probability distribution $f_a(\theta)$.

The most significant aspect of Bayesian estimation of unknown parameter θ is the determining of a certain function of a priori density probability distribution of the random variable $f_a(\theta)$. This function should correspond with preliminary information on the form of a priori density probability distribution of the random variable Θ. Here there are two issues. On the one hand, this information includes only a priori data. Therefore, the function of a priori density of probability distribution $f_a(\theta)$ should be selected based on the requirement of maximum possible entropy (in the sense of Shannon's theory [25]) under given conditions as specific a priori data being considered as limitations [18]. On the other hand, ignoring any objective a prior information can lead to choosing less informative function of a priori density of probability distribution for a random variable, which makes statistical estimation less effective.

In [32] the authors propose to enter the function of a priori density of probability distribution for a random variable with a defined parameter θ, which is considered as an unknown constant. Such a priori density of probability distribution gives a function of $f_a(\theta)$ only the a priori data, and at the same time allows using a priori objective information about the type of the law of distribution of the estimated parameter. Detailed calculations of the algorithm to specify parameter assessment θ shown in [16], and here we give only a general view of corresponding formula:

$$\hat{\theta} = \frac{\sum_{j=1}^{n} x_j f(x_j \,|\, \hat{\theta}') f_a(x_j \,|\, \hat{\theta}')}{\sum_{j=1}^{n} f(x_j \,|\, \hat{\theta}') f_a(x_j \,|\, \hat{\theta}')} . \tag{4}$$

Dependence available in this algorithm $\hat{\theta} = \varphi(x_j, j = \overline{1,n}; \hat{\theta}')$ can be solved by iteration [21], which implies the implementation of the following recurrent formula: $\hat{\theta}^{(l)} = \varphi(x_j, j = \overline{1,n}; \hat{\theta}^{(l-1)}), l \in [1, L]$, where l is a number of the current iteration, L is a number of iterations. Iterative procedure is completed after fulfilling the following condition: $\hat{\theta}^{(L)} - \hat{\theta}^{(L-1)} \le \lambda_\theta$ where λ_θ is defined calculation accuracy of assessment $\hat{\theta}$ of unknown parameter θ of a priori density of probability distribution of the random variable. To analyse the convergence we use the known theorem [21], according to which it is sufficient to follow the inequality provided above on the considered interval of the estimation $\hat{\theta}$ specifying for convergence of iterative procedure: $\left| d\varphi(x_j, j = \overline{1,n}; \hat{\theta}) / d\hat{\theta} \right| < 1$.

Thus, the most important aspect of Bayesian estimation of the unknown parameter is to assign a certain function of a priori density of probability distribution for a random variable. This function should correspond with the available preliminary information on the form of a priori probability distribution of this value.

Identifying of the most informative data channel. In complex synergistic systems, data on the state of the object, process or phenomenon is typically transmitted via several channels of communication [25], [31]. The problem is to determine which channels transmit the most reliable data. Therefore, the data received should be combined (generalised) to produce a cohered solution concerning the actual state of the object, process or phenomenon.

The traditional approach involves selecting of one or some of the most informative data channels and cutting off other less reliable channels. This is carried out by so-called mechanism of "reducers of degrees of freedom" described in [25]. The advantage of this method is either in the simplicity of implementation, or in its being the only possible and physically justified in many cases. However, some useful information that will not be involved into the process of developing of the coherent solution may be available in the cut-off communication channels.

To understand the information mentioned above, we give a known instructive example [8]. The ancient Chinese thinkers believed that any choice of one of several possible options was unjustified and unprofitable because it rejected all the others. The main concept of Chinese culture has always been a choice where the possible branches of the development of events were not cut off, but joined into a single whole, which was collectively acknowledged to be the only correct decision.

Therefore, during the synthesis of complex synergistic system in the process of data averaging we should ignore the concept of dominance of one of the data channels. This means that, instead of *reducers of degrees of freedom* we should enter the mechanisms that allow all data channels to participate in producing of solution with the significance corresponding to the degree of their information content in the current situation. Consequently, the information available will be properly used through the so-called *discriminators of degrees of freedom*.

Synergetic principle of data averaging has much in common with the ideas of mathematical statistics [29]. If the synergetic concept of data combination is used for evaluation of the actual state of the object, process or phenomena with the available set of data, mathematical statistics studies the methods of evaluating the moments of the probability distribution of random variables by the available set of sample elements. The common problems of both theories make the objectives of the research of synergistic features of mathematical statistics relevant for both synergy and development of statistical methods for data processing.

Let us consider the task of specified statistical estimation of mathematical expectation m_x of random variable X, distributed by the normal law with a posterior density $f(x|\theta) = f(x|m_x)$, given by the formula (2), according to the results of a random sample (1), if dispersion σ^2 is known. A priori information contains the estimation of mathematical expectation X_s, also distributed by the normal law of the known variance σ^2/n.

Having completed simple mathematical transformations described in [31], we obtained from expression (4) the algorithm for estimation of mathematical expectation X_s for its implementation we should organize the following iterative procedure:

$$X_s^{(l)} = \frac{\sum_{j=1}^{n} x_j Exp\left(-\frac{(x_j - X_s^{(l-1)})^2 (n+1)}{2\sigma^2} \right)}{\sum_{j=1}^{n} Exp\left(-\frac{(x_j - X_s^{(l-1)})^2 (n+1)}{2\sigma^2} \right)}, \forall l = 1, 2, 3, \ldots, \tag{5}$$

where l is a number of the current iteration. It is advisable to take maximum likelihood estimation (3) as zero approximation in this procedure, namely: $X_s^{(0)} = \frac{1}{n} \sum_{j=1}^{n} x_j$. Iterative procedure (5) during its application to problems of data averaging (examples 1–4) shows the mechanism for implementation of *discriminators of degrees of freedom*.

The results of numerous tests [8], [31], [16] show that the confidence interval that corresponds to the refined assessment (5) is shorter than the confidence interval of the maximum likelihood estimation (3). The largest gain in efficiency is obtained for small samples because the relative contribution of a prior information in obtaining estimates becomes smaller gradually with the increase in number of measurements, and Bayesian estimation (5) and maximum likelihood estimate (3) asymptotically coincide [29]. Therefore, it is expedient to calculate refined estimates of the parameters mainly for small volumes of random variable sample.

An important feature of a priori density of probability distribution of the random variable is that it should not be its own density, i.e. its integral should not necessarily be equal to one [2]. In some cases, the attempts to use pseudo Bayesian estimation are considered to be quite justified, when different density is introduced instead of missing a priori density of probability distribution of the random variable. Particularly noteworthy is the possibility to use it to show a priori density of probability distribution of the random variable for one of the so-called potential functions [1], the normal law of distribution being their special case (2). The following functions can also serve as alternatives:

$$f_1(\alpha)=\frac{\alpha}{|X_s-x|},\ f_2(\beta)=\frac{\beta}{(X_s-x)^2},\ f_3(\gamma,\delta)=\frac{\gamma}{1+\delta(X_s-x)^2}, \quad (6)$$

where α, β, γ, δ are some constant or current values of the random variable. Any of the potential function (6) is characterized by monotonically decrease with distance from the X_s value, that is symmetrical pair to evaluate X_s. If the estimated parameter is known to be distributed symmetrically in the general set, it is advisable to get refined estimation X_s, having selected quite a simple potential function for a priori density of probability distribution of the random variable.

Sometimes in order to reduce the amount of calculation it is advisable to replace the purposely known (e.g., normal) distribution law by another, simpler potential function [8]. For example, if a random variable is distributed by the uniform law, the assessment of its average value is a subject to the normal distribution law. However, if as the a priori density of probability distribution of the random variable we select function $f_1(\alpha)$ from the potential functions (6), we will arrive at a simple iterative algorithm for calculating mathematical expectation estimation X_s:

$$X_s^{(0)}=\frac{1}{n}\sum_{j=1}^{n}x_j;\ X_s^{(l)}=\sum_{j=1}^{n}\frac{x_j}{\left|X_s^{(l-1)}-x_j\right|\sum_{j=1}^{n}\frac{1}{\left|X_s^{(l-1)}-x_j\right|}}; \quad (7)$$
$$l\in[1,L],\ \left|X_s^{(l)}-X_s^{(l-1)}\right|\le\lambda_x.$$

This algorithm reflects the mechanism of implementation *of reducers of degrees of freedom* in its application to the procedure of solving the problem of data averaging. The above method of data averaging obtained from a limited amount of statistical material provides the individual approach to each implementation of the random variable – weighing under posteriori probability of its occurrence. That is, it allows eliminating the loss of valuable information when calculating the unknown parameter estimates from small samples [18].

If we select as the a priori density of probability distribution of the random variable function $f_2(\beta)$ or $f_3(\gamma,\delta)$ of the potential functions (6), we will arrive at quite different simple iterative algorithms for calculating a reliable estimate of mathematical expectation X_s, namely:

$$X_s^{(0)}=\frac{1}{n}\sum_{j=1}^{n}x_j;\ X_s^{(l)}=\sum_{j=1}^{n}\frac{x_j}{(X_s^{(l-1)}-x_j)^2\sum_{j=1}^{n}\frac{1}{(X_s^{(l-1)}-x_j)^2}}; \quad (8)$$
$$l\in[1,L],\ \left|X_s^{(l)}-X_s^{(l-1)}\right|\le\lambda_x,$$

$$X_s^{(l)}=\sum_{j=1}^{n}\frac{x_j}{(1+x_j(X_s^{(l-1)}-x_j)^2)\sum_{j=1}^{n}\frac{1}{1+x_j(X_s^{(l-1)}-x_j)^2}}; \quad (9)$$
$$l\in[1,L],\left|X_s^{(l)}-X_s^{(l-1)}\right|\le\lambda_x.$$

The author of [8] believes that for obtaining reliable estimate of mathematical expectation X_s through the organization of iterative procedures (8) and (9) all the elements of the sample of each iteration should interact. Synergistic system of data averaging similarly provides the procedure characterized by self-organization and self-management according to the purpose set. Here complex processes are developed through collective interaction of the components. Sequencing of component enables using reserve capacity of the synergetic system and thus significantly increases the emergence degree (systemic effect).

Thus, the mechanism of reducers of degrees of freedom is the traditional approach to identify the most informative channel of transferring data on the state of the object, process or phenomena and cutting off others, less reliable data [34]. Its main drawback is that communication channels cut-off may contain some useful information that will not be involved in the process of developing a cohered solution. Therefore, it is necessary to introduce the mechanisms of discriminators of degrees of freedom that allow all data channels to participate in developing a cohered solution with the significance that corresponds to the largest degree of their information content in the current situation.

3. An illustrative example of the data averaging mechanism

Let the number of data channels be $n \ge 3$ (the number of degrees of freedom in the synergistic system of data averaging). We present input as a one-dimensional array $\tilde{X}=\{x_j,\ j=\overline{1,n}\}$, where $x_j,\ j=\overline{1,n}$ is the value of some random variable X, obtained through j-th channel of communication (the component of the data averaging system).

Depending on the practical requirements (see Examples 1-4), we should identify the most informative data channel and a reliable estimate x^* of the random variable X. Let the input array be as follows: $\tilde{X}=\{x_j,\ j=\overline{1,5}\}$ = {5.0, 6.5, 4.3, 5.2, 6.0}. The condition for stopping iterative procedure is $\left|X_s^{(l)}-X_s^{(l-1)}\right|\le\lambda_x$ = 0.005. Estimate of zero iteration is as follows: $X_s^{(0)}$ = 1/5·(5.0+6.5+4.3+5.2+6.0) = 27.0/5 = 5.40.

Option 1. Let us consider the problem of determining the most informative data channel and apply the mechanism for implementation of the reducers of the degrees of freedom expressed by the iterative algorithm (7). The results of calculations by iteration are given in Table 2.

Thus, as a result of the iterative algorithm implementation (7) we received a reliable estimate of the random variable that is $x^*=X_s^{(l^*)}=X_s^{(5)}=5.20$, and the most informative data channel should be determined by the following expression: $x^*=\min\left\{\left|X_s^{(l^*)}-x_j\right|,\ j=\overline{1,n}\right\}$. In our example, the most informative appeared to be the fourth channel. The figure 1 shows the corresponding results of calculations performed using appropriate software application.

Option 2. Let us consider the problem of determining the most accurate assessment of a random variable X_s through the mechanism for implementation of discriminators of degrees of freedom expressed by the iterative algorithm (5). It is assumed that the random variable X_s is distributed in the general set normally with known dispersion. The results of calculation by iteration are shown in Table 3.

Table 2. Results of calculations by iteration using the mechanism for implementation of the reducers of the degrees of freedom

Iteration	The experts and their assessment:					Sum	Marking	Average value	$\lambda_x =$	0.005
	1	2	3	4	5				$\lvert X_c[l]-X_c[l-1]\rvert$	
	5.0	**6.5**	**4.3**	**5.2**	**6.0**	27.000	$X_s[0]$	5.400		
1	0.400	1.100	1.100	0.200	0.600					
	2.500	0.909	0.909	5.000	1.667	10.985				
	1.138	0.538	0.356	2.367	0.910	5.309	$X_s[1]$	5.309	0.091	failed
2	0.309	1.191	1.009	0.109	0.691					
	3.237	0.840	0.991	9.177	1.447	15.692				
	1.031	0.348	0.272	3.041	0.553	5.245	$X_s[2]$	5.245	0.064	failed
3	0.245	1.255	0.945	0.045	0.755					
	4.078	0.797	1.058	22.106	1.325	29.363				
	0.694	0.176	0.155	3.915	0.271	5.211	$X_s[3]$	5.211	0.034	failed
4	0.211	1.289	0.911	0.011	0.789					
	4.735	0.776	1.097	89.437	1.268	97.314				
	0.243	0.052	0.048	4.779	0.078	5.201	$X_s[4]$	5.201	0.010	failed
5	0.201	1.299	0.901	0.001	0.799					
	4.977	0.770	1.110	1105.01	1.251	1113.12				
	0.022	0.004	0.004	5.162	0.007	5.200	$X_s[5]$	**5.200**	0.001	done
	0.200	1.300	0.900	*0.000*	0.800	*0.000*	← min			

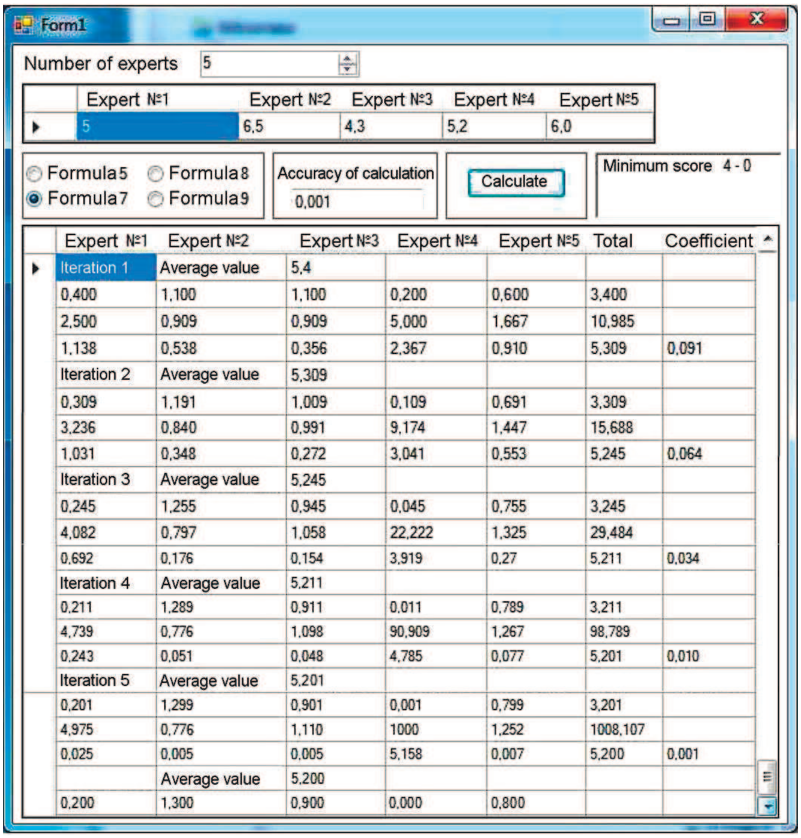

Figure 1. Software application for performing relevant calculations

Thus, we have obtained the value of the random variable $x^* = X_c^{(l^*)} = X_c^{(6)} = 5.081$, that is the most reliable estimates under given assumptions. In our example, the most informative appeared to be the first channel. The figure 2 shows the corresponding results of calculations performed using appropriate software application.

The conducted above comparative analysis of the features of the implementation of iterative algorithms (5) and (7), which are typical both for the methods of mathematical statistics and methods of synergetic system of data averaging showed the advantages of some of them and disadvantages of others. The results of the analysis can be used to improve the reliability of obtaining statistical estimation of the random variable which is calculated in the presence of a small sample of expert based evaluation or experimental data, and also to determine assessments of the actual state of the object, the procedure or phenomenon in synergetic systems of data averaging in the limited number of channels.

Table 3. Results of calculations by iteration using the mechanism for implementation of discriminators of degrees of freedom

Itera-tion	The experts and their assessment:					Sum	Marking	Average value	λ_x =	0.005
	1	2	3	4	5				$\lvert X_c[l]-X_c[l-1]\rvert$	
	5.0	**6.5**	**4.3**	**5.2**	**6.0**	27.000	$X_s[0]$	5.400		
1	0.644	4872	4.872	0.161	1.450					
	0.525	0.008	0.008	0.851	0.235	1.626				
	2.625	0.050	0.033	4.426	1.408	8.542	$X_s[1]$	5.253	0.147	failed
2	0.257	6.264	3.655	0.011	2.249					
	0.773	0.002	0.026	0.989	0.106	1.895				
	3.866	0.012	0.111	5.142	0.633	9.765	$X_s[2]$	5.152	0.101	failed
3	0.093	7.317	2.923	0.009	2.896					
	0.911	0.001	0.054	0.991	0.055	2.012				
	4.556	0.004	0.231	5.152	0.332	10.275	$X_s[3]$	5.108	0.044	failed
4	0.047	7.805	2.627	0.034	3.206					
	0.954	0.000	0.072	0.966	0.041	2,034				
	4.772	0.003	0.311	5.025	0.243	10.353	$X_s[4]$	5.090	0.017	failed
5	0.033	8.001	2.516	0.048	3.332					
	0.968	0.000	0.081	0.953	0.036	2.037				
	4.838	0.002	0.348	4.954	0.214	10.357	$X_s[5]$	5.084	0.007	failed
6	0.028	8.079	2.472	0.055	3.382					
	0.972	0.000	0.084	0.947	0.034	2.038				
	4.861	0.002	0.363	4.924	0.204	10.354	$X_s[5]$	**5.081**	0.003	done
	0.081	1.419	0.781	0.119	0.919	*0.081*	← min			

Note: n = 5; $2\sigma^2$ = 1.490; $(n+1)/2\sigma^2$ = 4.027.

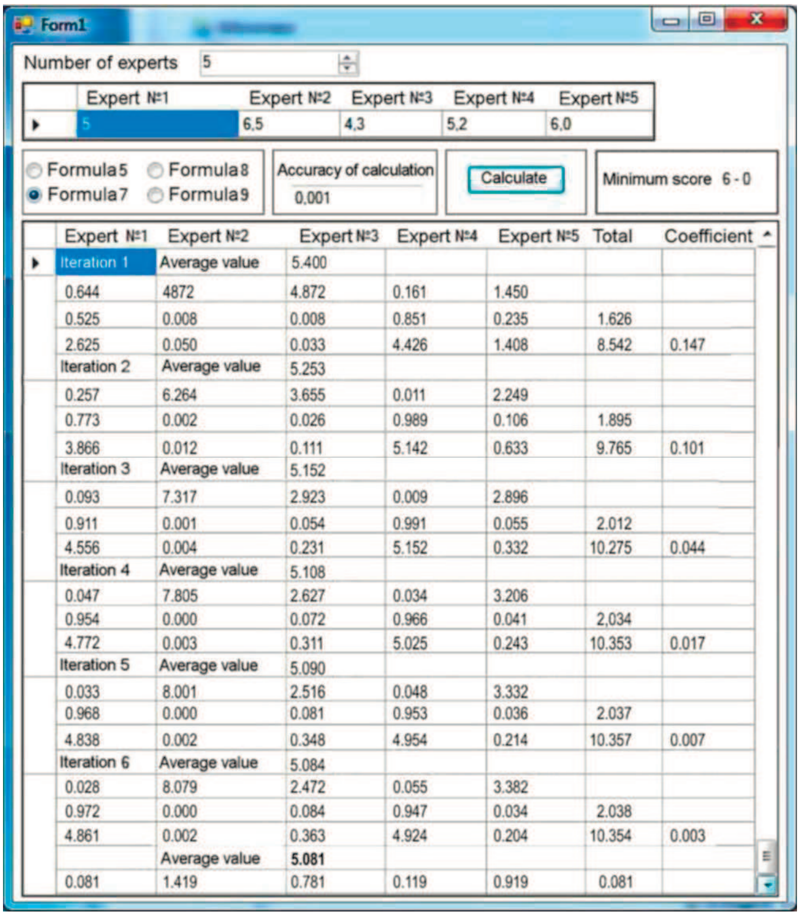

Figure 2. The results of the calculations made using the appropriate software application

Conclusions. The methods for specified point estimates of the parameter of the probability distribution of the random variable based on the limited amount of statistical data, which helped identify the most informative data channel and get it accurate assessment. Having conducted the research, we have drawn the following conclusions.

1. We have provided some examples that demonstrate the need to determine the relative degree of reliability of data obtained by different communication channels on the same state of the object, process or phenomenon and the need of averaging. We have also revealed that data analyses and processing are carried out with the involvement of known techniques of probability theory and mathematical statistics, where considerable theoretical and practical experience is accumulated. The model, which describes the state of the object, process or phenomenon presented as point estimates

of the parameters of probability distribution of the random variable which value is obtained based on a small sample, is developed.

2. Modern approaches to evaluating statistical random variable, the most common of which is the Bayesian approach, are analysed. We have defined that the most important aspect of Bayesian estimation of the unknown parameter is to assign a priori function of probability density of the random variable. This function should correspond with preliminary information on the form of a priori probability distribution of this value.

3. We have considered the traditional approach to identifying the most informative data channel on the state of the object, process or phenomena and cutting off other less reliable data channels. This is performed by so-called mechanism of reducers of degrees of freedom, which main drawback is that the cut off communication channels may contain some useful information, which is not involved in the process of developing the coherent solution. Therefore, we suggest introducing mechanisms of discriminators of degrees of freedom that allow all data channels to participate in the development of solutions with significance that corresponds to the degree of their information content in the current situation.

4. We have provided the illustrative example of the data averaging mechanism that shows the results of calculations by iterations using mechanisms of implementation for both reducers and discriminators of degrees of freedom. These arrangements reflect the peculiarities of the implementation of iterative algorithms that are characteristic both for specific methods of mathematical statistics and methods for synergetic system of data averaging.

5. We have made the appropriate conclusions and recommendations regarding the use of the developed methods for specified point estimates of the parameter of the probability distribution of the random variable based on the limited amount of statistical material.

References

[1] Aizerman, M. A., Braverman, E. M., & Rozonoer, L. I. (1970). The method of potential functions in machine learning theory. Moscow: Science. 384 p. [In Russian].

[2] Bakhrushin, V. E. (2006). *Data Analysis: a tutorial*. Zaporizhzhia: PG "Humanities", 128 p. [In Ukrainian].

[3] Bakhrushin, V. E., & Ignahina, M. A. (2008). Application of statistical methods in processing the results of production control in metallurgy of semiconductors. *System Technology, 3*(56), Vol. 1, 3–7. [In Russian].

[4] Botsula, M., & Morgun, I. (2008). The problem of quality examination of distance courses. *Scientifical Journals of Vinnytsia National Technical University, 4*, 1–7. Retrieved from: http://nbuv.gov.ua/e-iournals/vntu/2008-4/2008-4.files/uk/08mpbcme.uk.pdf. [In Ukrainian].

[5] Brandt, Z. (2003). Data analysis: Statistical and Computational Methods for Scientists and Engineers. Moscow: Mir, AST, 686 p. [In Russian].

[6] Gaskarov, D., & Shapovalov, V. I. (1978). *Small sample*. Moscow: Statistics, 248 p. [In Russian].

[7] Gmurman, B. E. (2004). Guide to solving problems of the theory of probability and mathematical statistics. Moscow: Higher School, 404 p. [In Russian].

[8] Gmurman, V. E. (2003). *Probability theory and mathematical statistics*. Moscow: Higher School, 479 p. [In Russian].

[9] Gryciuk, Yu. I., & Grytsyuk, P. Yu. (2019). Contemporary problems of scientific evaluation of the application software quality. *Scientific Bulletin of UNFU, 25*(7), 284–294. https://doi.org/10.15421/40250745

[10] Guter, R. S., & Reznikovskii, P. T. (1971). *Programming and computational mathematics*. Moscow: Science. Vol. 2, 273 p. [In Russian].

[11] Hrytsiuk, Yu. I., & Andrushchakevych, O. T. (2018). Means for determining software quality by metric analysis methods. *Scientific Bulletin of UNFU, 28*(6), 159–171. https://doi.org/10.15421/40280631.

[12] Hrytsiuk, Yu. I., & Buchkovska, A. Yu. (2017). Visualization of the Results of Expert Evaluation of Software Quality Using Polar Diagrams. *Scientific Bulletin of UNFU, 27*(10), 137–145. https://doi.org/10.15421/40271025

[13] Hrytsiuk, Yu. I., & Grytsyuk, P. Yu. (2019). The methods of the specified points of the estimates of the parameter of probability distribution of the random variable based on a limited amount of data. *Scientific Bulletin of UNFU, 29*(2), 141–149. https://doi.org/10.15421/40290229

[14] Hrytsiuk, Yu. I., & Nemova, E. A. (2018). Peculiarities of Formulation of Requirements to the Software. *Scientific Bulletin of UNFU, 28*(7), 135–148. https://doi.org/10.15421/40280727.

[15] Kartavy, V., & Yarovaya, V. (2004). *Mathematical Statistics*. Kyiv: Professional, 484 p. [In Ukrainian].

[16] Kobzar, A. I. (2006). *Applied Mathematical Statistics*. Moscow: Fizmatlit, 816 p. [In Russian].

[17] Kolesnikov, A. A. (1994). *Synergetic theory of management*. Moscow: Energoatomisdat, 344 p. [In Russian].

[18] Lagutin, M. B. (2007). *Transparent mathematical statistics*. Moscow: Binom, 472 p. [In Russian].

[19] Mikitin, J. P. (2008). Programming model averaging method Noise. Bulletin of the National University "Lviv Polytechnic". Series: Computer Science and Information Technology, 629, 21–24. [In Ukrainian].

[20] Morgun, I. (2011). The method of peer review software quality. Software Engineering: mater. *Intern. nauk. and practical. Conf. graduate students, 2*(6), 33–37. Vinnytsia. Retrieved from: http://jrnl.nau.edu.ua/index.php/IPZ/article/view/3086. [In Ukrainian].

[21] Orlov, A. I. (2006). *Applied Statistics*. Moscow: Exam, 671 p. [In Russian].

[22] Pleskach, V. L., & Zatonatska, T. (2011). *Information systems and technology in enterprises: textbook*. Kyiv: Knowledge. 718 p. Retrieved from: http://pidruchniki.com/1194121347734/informatika/analiz_yakosti_programnogo_zabezpechennya#42. [In Ukrainian].

[23] Protasov, K. V. (2005). *Statistical analysis of experimental data*. Moscow: Mir, 142 p. [In Russian].

[24] Sage, E., & Mels, J. (1976). Estimation theory and its application in communication and management. Moscow: Communication, 496 p. [In Russian].

[25] Shannon, K. (1963). *Work on information theory and cybernetics*. Moscow: Publishing House of Foreign Literature, 829 p. [In Russian].

[26] Tolbatov, A. (1994). Mathematical Statistics and task optimization algorithms and programs. Kyiv: High School, 226 p. [In Ukrainian].

[27] Tuluchenko, G. Y. (2008). Geometry computing templates bars of centric averaging method. *Bulletin of the Zaporizhzhya National University, 1*, 187–190. [In Ukrainian].

[28] Turchin, V. (2006). *Probability and Mathematical Statistics: Concepts, examples, problem*. Dnepropetrovsk: Dniprovsky National University, 476 p. [In Ukrainian].

[29] Vankovych, T.-N. M., Zinko, J. A., & Bozhenko, M. (2010). An averaging method for oscillating stochastic systems with quick phase. *Bulletin of the National University "Lviv Polytechnic". Series: Dynamics, Durability and Design of Machines and Devices, 678*, 11–14. [In Ukrainian].

[30] Voronin, A. N. (2004). Method of interconnecting signals for bistatic radar small celestial bodies. *System analysis and ma-*

nagement: meas. rep. 9th International. Conf., (pp. 113–114). Moscow: Publishing house of the Moscow Aviation Institute. [In Russian].

[31] Voronin, A. N. (2006). Synergistic methods of data aggregation. *Cybernetics and Systems Analysis, 2*, 24–30. [In Russian].

[32] Voronin, A. N. (2014). Methods of data aggregation. *Cybernetics and Systems Analysis, 50*(5), 78–84. [In Russian].

[33] Voronin, A. N., & Ziatdinov, J. K. (2013). *Theory and practice of multi-criteria decisions: models, methods, implementation*. Saarbrucken (Deutschland); Lambert Academic Publishing, 305 p. [In Russian].

[34] Voronin, A. N., Ziatdinov, J. K., & Kulinskiy, M. V. (2011). *Multicriteria task: models and methods: a monograph*. Kyiv: NAU, 348 p. [In Russian].

[35] Zhluktenko, V. I., & Nakonechny, S. (2000). *Probability and Mathematical Statistics: training method. manual.* In 2 parts. Part I. Probability. Kyiv: Kyiv National Economic University, 304 p. [In Ukrainian].

[36] Zhluktenko, V. I., Nakonechny, S., & Savin, S. (2001). *Probability and Mathematical Statistics: training method. manual*. In 2 parts. Part II. Mathematical Statistics. Kyiv: Kyiv National Economic University, 336 p. [In Ukrainian].

Ю. І. Грицюк, О. Р. Ферneза

Національний університет "Львівська політехніка", м. Львів, Україна

УТОЧНЕННЯ ЕКСПЕРТНИХ ОЦІНОК НА ПІДСТАВІ ОБМЕЖЕНОГО ОБСЯГУ ДАНИХ

Розроблено методику уточнених експертних оцінок параметра розподілу ймовірностей випадкової величини на підставі обмеженого обсягу статистичних даних. Це дало змогу виявити найбільш інформативний канал передачі даних (кваліфікованого експерта) і отримати його достовірну оцінку. Встановлено, що аналіз та оброблення даних здійснюють із залученням відомих методик з теорії ймовірностей та математичної статистики, де нагромаджено значний теоретичний і практичний досвід. Математичну модель, яка описує стан деякого об'єкта, процесу чи явища, подано у вигляді точкової оцінки параметра розподілу ймовірностей випадкової величини, значення якого отримують на підставі малої вибірки. Проаналізовано сучасні підходи до статистичного оцінювання випадкової величини, найпоширенішим з яких є Байєсовський підхід. Встановлено, що найбільш значущим моментом Байєсового оцінювання невідомого параметра є призначення певної функції апріорної щільності розподілу ймовірностей випадкової величини. Ця функція має відповідати наявній попередній інформації про форму апріорного розподілу ймовірностей цієї величини.

Розглянуто традиційний підхід до виявлення найбільш інформативного каналу передачі даних про стан об'єкта, перебіг процесу чи явища і відсікання інших – менш достовірних. Це здійснюють за допомогою так званого механізму редукторів ступенів свободи. Його основний недолік полягає в тому, що у відсічених каналах зв'язку може існувати деяка корисна інформація, яка не бере участі в процесі вироблення узгодженого рішення. Тому потрібно вводити механізми дискримінаторів ступенів свободи. Вони дадуть змогу всім каналам передачі даних брати участь в процесі підготовки рішення з вагомістю, яка відповідає найбільшому ступеню їх інформативності в поточній ситуації. Наведено ілюстративний приклад застосування розглянутих методів усереднення даних, у якому відображено результати розрахунків за ітераціями з використанням механізмів реалізації як редукторів, так і дискримінаторів ступенів свободи. Ці механізми відображають особливості реалізації ітераційних алгоритмів, характерних як для методів математичної статистики, так і для методів синергетичної системи усереднення даних.

Ключові слова: теорія ймовірностей; математична статистика; методи усереднення даних; інформативний канал передачі даних; механізм редукторів ступенів свободи; механізм дискримінаторів ступенів свободи; ітераційні алгоритми.

Інформація про авторів:

Грицюк Юрій Іванович, д-р техн. наук, професор кафедри програмного забезпечення. **Email:** yurii.i.hrytsiuk@lpnu.ua; https://orcid.org/0000-0001-8183-3466; ResearcherID: V-3995-2017

Ферneза Остап Романович, аспірант, кафедра програмного забезпечення. **Email:** ostapfer@gmail.com

Цитування за ДСТУ: Грицюк Ю. І., Ферneза О. Р. Уточнення експертних оцінок на підставі обмеженого обсягу даних. *Український журнал інформаційних технологій.* 2019, т. 1, № 1. С. 58–66.

Citation APA: Hrytsiuk, Yu. I., & Ferneza, O. R. (2019). Refining expert based evaluation on the basis of a limited quantity of data. *Ukrainian Journal of Information Technology, 1*(1), 58–66. https://doi.org/10.23939/ujit2019.01.058