# Information systems research

UDC 519.2 **doi:** https://doi.org/10.20998/2522-9052.2023.1.05

Lev Raskin[1], Larysa Sukhomlyn[2], Dmytro Sokolov[1], Vitalii Vlasenko[1]

[1] National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine
[2] Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine

## EVALUATION OF SYSTEM CONTROLLED PARAMETERS INFORMATIONAL IMPORTANCE, TAKING INTO ACCOUNT THE SOURCE DATA INACCURACY

**Abstract.** The work considers system controlled parameters information value assessing technology in the task of its state identifying. The purpose of the study is to improve the standard methodology for controlled parameters information value assessing. The proposed method is based on the controlled parameters value probabilities analysis falling into the subintervals of the interval of possible values for different states of the system. When the value of the controlled parameter falls into the left or right boundary subintervals of the compatibility interval for any state of the object, the conclusion about its state is made taking into account possible errors of the first or second kind in this case. When the controlled parameter value enters the central subinterval, useful information appears if the corresponding probabilities for the states $H_1$ and $H_2$ are differ significantly. Thus, it is shown that taking into account the probabilities of fuzzy values of the controlled parameter falling into the compatibility interval for various states of the object significantly increases its informational value.

**Keywords:** task of system states identifying; information value of the controlled parameters; taking into account the probabilities of their falling into the intervals of possible values.

### Introduction

**Problem Statement.** Task of object state identifying is an integral component of technologies for solving many problems of systems effectiveness evaluating and their optimal management. Information base for solving these problems is a set of measurement results of system controlled parameters values, the evolution of which is somehow connected with system state evolution. Special statistical methods are used to solve the problem of system state identifying. Consider the essence of the most practically frequently used.

The method of multidimensional discriminant analysis [1] in the most commonly used case solves the problem of state identification as follows. It is assumed that the system can be in one of two states: $H_1$ or $H_2$. To assess the state, a set of controlled parameters $(x_1, x_2, ..., x_p)$ is used, each of which is a random variable with a known mathematical expectation $m_{1i}$, if the system is in a state $H_1$, and $m_{2i}$ if the state of the system is $H_2$, $i = 1, 2, ..., p$. In addition, it is assumed that the matrix of correlation coefficients values between the parameters $K = (k_{i1,i2}),\ i_1 = 1, 2, .., p,\ i_2 = 1, 2, .., p.$ ++is known. The initial statistical information about the controlled parameters is used to calculate the discriminant function

$$D(x) = \sum_{i=1}^{p} a_i x_i, \tag{1}$$

the coefficients of which are determined by equation system solution

$$\sum_{i=1}^{p} k_{si}\ a_i = m_{1s} - m_{2s}, s = 1, 2, ..., p. \tag{2}$$

The set $(a_1, a_2, ..a_p)$ obtained as a result of solving the equations system (2) is used for calculation

$$\varsigma_1 = \sum_{i=1}^{p} a_i m_{1i},\ \ \varsigma_2 = \sum_{i=1}^{m} a_i m_{2i}, C = \frac{1}{2}(\varsigma_1 + \varsigma_2). \tag{3}$$

The calculated values $C$ allow to formulate an identification rule: for a specific set of values of the controlled parameters $x = (x_1, x_2, .., x_p)$ the system is in a state $H_1$ if the value of the discriminant function on this set satisfies the inequality

$$D(x) = \sum_{i=1}^{m} a_i x_i < C, \tag{4}$$

and is in a state $H_2$ otherwise.

At the same time, as shown in [1,2], the above rule provides a minimum of error total probability in object state assessing.

Let's list the obvious disadvantages of the described technology:

– the method of solving the identification problem presented by the relations (1)–(4) is simply implemented if the set of possible states of the system consists of two states; the task becomes significantly more complicated if the number of states is more than two;

– the theoretical justification of the method is based on the hypothesis of controlled parameters values normality; in real conditions of a small sample of initial data, this hypothesis cannot be correctly confirmed or rejected.

In this regard, the application of this method in practical conditions can lead to gross errors in the condition identification.

A more reliable procedure for identifying the system, free from limiting requirements for the dimension of the problem, is cluster analysis [3, 4].

Let the system be in one of the many $(H_1, H_2, .., H_K)$ possible states at any given time. For each, for example, $k$ state, a test set of controlled system parameters values corresponding to this state –

$(m_{k1}, m_{k2}, .., m_{kp})$ is formed. This set in $p$ – dimensional coordinates defines a point $A_k$, $k = 1, 2, .., K$ .These points form a set of grouping centers. The clustering procedure is implemented as follows. Suppose, as a result of measuring the appropriate values, a set of $x = (x_1, x_2, .., x_p)$ values of the controlled parameters is obtained, defining a specific point in $p$ – dimensional parameter space. For this point, the distances to each of the grouping centers are calculated in the selected metric, for example, Euclidean:

$$R_k = \sum_{i=1}^{p} (x_i - m_{ki})^2, k = 1, 2, ..., K.$$

The minimum of these distances determines the nearest of the grouping centers that define specific states. For each new measurement of a set of controlled parameters, the described procedure for attaching this new point to one of the grouping centers is repeated. Analysis of the results of a certain set of measurements allows us to conclude about the possible state of the system.

If for some reason the position of the grouping centers is not set a priori, then, using the known procedures [3,4], the problem of these centers coordinates calculating is successfully solved.

It is clear that the accuracy of solving the problem of system state identifying depends significantly on how much the distribution densities of controlled parameters random values differ from each other for different states of the system, that is, on the degree of these parameters informativeness.

Constructive consideration of the measure of random values distribution densities proximity of the controlled parameter for different states of the system is implemented in the Kullback criterion [5]. For the special case when the set of system possible states contains only two states: $H_1$ and $H_2$ , the Kullback criterion is constructed as follows. Let $\varphi(x/H_1)$ and $\varphi(x/H_2)$ be the distribution densities of the controlled parameter $x$ for these states. Then the criterion values are calculated by the formula:

$$\tau = \int_{-\infty}^{\infty} \varphi_1(x/H_1) \log \frac{\varphi_1(x/H_1)}{\varphi_2(x/H_2)} dx. \quad (5)$$

If these distribution densities are Gaussian,then (5) takes the form:

$$\tau = \int_{-\infty}^{\infty} \left( 1 \Big/ \left( \sqrt{2\pi}\ \sigma_1 \right) \right) \exp\left\{ -(x - m_1)^2 \Big/ \left( 2\sigma_1^2 \right) \right\} \times$$
$$\times \log \frac{\left( \sqrt{2\pi}\sigma_1 \right)^{-1} \cdot \exp\left\{ -(x - m_1)^2 \Big/ \left( 2\sigma_1^2 \right) \right\}}{\left( \sqrt{2\pi}\sigma_2 \right)^{-1} \cdot \exp\left\{ -(x - m_2)^2 \Big/ \left( 2\sigma_2^2 \right) \right\}} dx. \quad (6)$$

The disadvantages of criterion (5) are obvious.

1) Numerical value of criterion (5) is 0 if the densities of $\varphi_1(x/H_1)$ and $\varphi_2(x/H_2)$ coincide. Otherwise, measure (5) takes a positive, non-top-bounded value. At the same time, the calculated value of measure (5) in each specific case does not contain clear, easily interpreted information about whether the controlled parameter is sufficiently informative.

2) A significantly more serious disadvantage of the Kullback information measure is its asymmetry, that is, the result of measure calculation depends on the nature (method) of the densities $\varphi_1(x/H_1)$ and $\varphi_2(x/H_2)$ entering into the calculation formula (5). At the same time, the difference in the numerical value of the measure for the relations

$$\tau_1 = \int_{-\infty}^{\infty} \varphi_1(x/H_1) \log\left( \varphi_1(x/H_1) / \varphi_2(x/H_2) \right) dx,$$
$$\tau_2 = \int_{-\infty}^{\infty} \varphi_2(x/H_2) \log\left( \varphi_2(x/H_2) / \varphi_1(x/H_1) \right) dx.$$

can be unpredictably large.

These circumstances motivate the consideration and selection of other options for assessing the degree of controlled parameters informativeness.

## Analysis of known results

In [6], an analysis of alternative methods for controlled parameters informativeness measure calculation was carried out and the following simple method for informativeness evaluation was proposed, which is implemented as follows. Parameter informativeness measure is introduced by the ratio:

$$\tau\left( \varphi_1(x/H_1), \varphi_2(x/H_2) \right) =$$
$$= 1 - \int_{-\infty}^{\infty} \sqrt{\varphi_1(x/H_1) \varphi_2(x/H_2)} dx. \quad (7)$$

If the densities $\varphi_1(x/H_1)$ and $\varphi_2(x/H_2)$ are Gaussian, then

$$\tau\left( \varphi_1(x/H_1), \varphi_2(x/H_2) \right) = 1 - \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{(x - m_1)^2}{2\sigma_1^2} \right\} \frac{1}{\sqrt{2\pi}\sigma_2} \times \exp\left\{ -\frac{(x - m_2)^2}{2\sigma_2^2} \right\} \right)^{\frac{1}{2}} dx = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}} \times$$
$$\times \exp\left\{ -\frac{1}{2} \left[ \frac{(x - m_1)^2}{2\sigma_1^2} + \frac{(x - m_2)^2}{2\sigma_2^2} \right] \right\} dx = = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}} \exp\left\{ -\frac{1}{4{\sigma_1}^2{\sigma_2}^2} \left[ x^2{\sigma_1}^2 + x^2{\sigma_2}^2 - 2xm_1{\sigma_2}^2 - 2xm_2{\sigma_1}^2 + \right. \right.$$
$$\left. \left. + {m_1}^2{\sigma_2}^2 + {m_2}^2{\sigma_1}^2 \right] \right\} dx = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi{\sigma_1}^2{\sigma_2}^2}} \times \exp\left\{ -\frac{{\sigma_1}^2 + {\sigma_2}^2}{4{\sigma_1}^2{\sigma_2}^2} \left[ x^2 - 2x \frac{m_1{\sigma_2}^2 + m_2{\sigma_1}^2}{{\sigma_1}^2 + {\sigma_2}^2} + \frac{{m_1}^2{\sigma_2}^2 + {m_2}^2{\sigma_1}^2}{{\sigma_1}^2 + {\sigma_2}^2} \right] \right\} dx =$$

$$=1-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_1{}^2\sigma_2{}^2}\exp\left\{-\frac{\sigma_1{}^2+\sigma_2{}^2}{4\sigma_1{}^2\sigma_2{}^2}\left[\left(x-\frac{m_1\ \sigma_2{}^2+m_2\ \sigma_1{}^2}{\sigma_1{}^2+\sigma_2{}^2}\right)^2-\left(\frac{m_1\ \sigma_2{}^2+m_2\ \sigma_1{}^2}{\sigma_1{}^2+\sigma_2{}^2}\right)+\frac{m_1{}^2\sigma_2{}^2+m_2{}^2\sigma_1{}^2}{\sigma_1{}^2+\sigma_2{}^2}\right]\right\}dx=$$

$$=1-\int_{-\infty}^{\infty}\sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1{}^2+\sigma_2{}^2}}\times\frac{\sqrt{\sigma_1{}^2+\sigma_2{}^2}}{\sqrt{2\pi}\cdot\sigma_1\sigma_2}\exp\left\{-\frac{\left(x-\left(m_1\sigma_2{}^2+m_2\sigma_2{}^2\right)\Big/\left(\sigma_1{}^2+\sigma_2{}^2\right)\right)}{2\sigma_1\sigma_2\Big/\left(\sigma_1{}^2+\sigma_2{}^2\right)}\right\}e^{-A}dx, \quad (8)$$

where $A=\left(4\left(\sigma_1{}^2+\sigma_2{}^2\right)\sigma_1{}^2\sigma_2{}^2\right)^{-1}\left(m_1{}^2\sigma_1{}^2\sigma_2{}^2+m_1{}^2\sigma_2{}^4+m_2{}^2\sigma_1{}^4+m_1{}^2\sigma_1{}^2\sigma_2{}^2-m_1{}^2\sigma_2{}^2-2m_1m_2\sigma_2{}^4\sigma_1{}^2-\right.$

$$\left.-m_2{}^2\sigma_1{}^4\right)==\left(m_1-m_2\right)^2\Big/\left(4\left(\sigma_1{}^2+\sigma_2{}^2\right)\right).$$

It follows that

$$0\le\tau\left(\varphi_1\left(x/H_1\right),\varphi_2\left(x/H_2\right)\right)<1,$$

moreover, equality $\tau=0$ is achieved only if $m_1=m_2$ and at the same time $\sigma_1=\sigma_2$.

The introduced measure (7) is symmetric and its value approaches unity as the discrepancy between $\varphi_1\left(x/H_1\right)$ and $\varphi_2\left(x/H_2\right)$ increases. The obvious disadvantage is the computational complexity of relation (7) in the general case.

In [6], another possible criterion for the informativeness of the parameters is considered, which is based on the study of plots areas arising at the intersection of densities $\varphi_1\left(x/H_1\right)$ and $\varphi_2\left(x/H_2\right)$. At the same time, the functions are introduced

$$\varphi_{1,2}\left(x\right)=\min\left\{\varphi_1\left(x/H_1\right),\varphi_2\left(x/H_2\right)\right\},$$

$$\breve{\varphi}_{1,2}\left(x\right)=\max\left\{\varphi_1\left(x/H_1\right),\varphi_2\left(x/H_2\right)\right\}.$$

Further it is necessary to calculate

$$S=\left(\varphi_{1,2}\left(x\right)\right)=\int_{-\infty}^{x^*}\varphi_2\left(x/H_2\right)dx+\int_{x^*}^{\infty}\varphi_1\left(x/H_1\right)dx, \quad (9)$$

$$\breve{S}=\left(\varphi_{1,2}\left(x\right)\right)=\int_{-\infty}^{x^*}\varphi_1\left(x/H_1\right)dx+\int_{x^*}^{\infty}\varphi_2\left(x/H_2\right)dx, \quad (10)$$

where $x^*$ is the intersection point of the curves $\varphi_1\left(x/H_1\right)$ and $\varphi_2\left(x/H_2\right)$.

At the same time (9) determines the area of densities intersection, and (10) is the area of their union. Now the informativeness measure of the parameter $x$ is given by the ratio

$$\tau=1-S\left(\varphi_{1,2}\left(x\right)\right)\Big/\breve{S}\left(\varphi_{1,2}\left(x\right)\right). \quad (11)$$

The value $\tau$ calculated in accordance with (11) is also zero if $S\left(\varphi_{1,2}\left(x\right)\right)=\breve{S}\left(\varphi_{1,2}\left(x\right)\right)$, that is, if $\varphi_1\left(x/H_1\right)=\varphi_2\left(x/H_2\right)$, and monotonically approaches unity as the area of intersection of the figures corresponding to $\varphi_1\left(x/H_1\right)$ and $\varphi_2\left(x/H_2\right)$ decreases. The computational complexity of criterion (11) calculation is determined by the complexity of solving the equation $\varphi_1\left(x/H_1\right)=\varphi_2\left(x/H_2\right)$, which is numerically possible in the general case. In this regard, the following is proposed, which naturally follows from (11), a simple way to assess the informative value of the indicator. Let, for example, the distribution densities of the controlled parameter be Gaussian. Select some sufficiently small number $\varepsilon$, for example $\varepsilon=0,05$, and two equations are solved

$$\left(\sqrt{2\pi}\sigma_1\right)^{-1}\cdot\exp\left\{-\left(x-m_1\right)^2\Big/\left(2\sigma_1{}^2\right)\right\}=\varepsilon, \quad (12)$$

$$\left(\sqrt{2\pi}\sigma_2\right)^{-1}\cdot\exp\left\{-\left(x-m_1\right)^2\Big/\left(2\sigma_1{}^2\right)\right\}=\varepsilon. \quad (13)$$

Let $\left(a_1,b_1\right)$ be the roots of equation (12), and let $\left(a_2,b_2\right)$ be the roots of equation (13). The resulting situation is shown in Fig. 1.

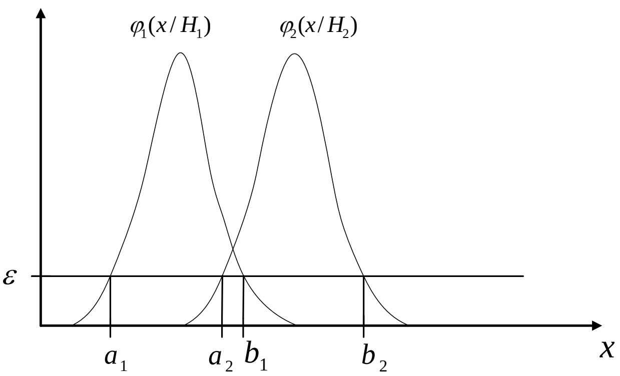


**Fig. 1.** Distribution densities $\varphi_1\left(x/H_1\right)$ and $\varphi_2\left(x/H_2\right)$

Now, with a sufficiently high probability determined by the value $\varepsilon$, it can be assumed that if the observed value $x$ of the controlled parameter belongs to an interval $x\in I_1=[a_1,a_2]$, then the system is in a state $H_1$, if $x\in I_2=[b_1,b_2]$, then the system is in a state $H_2$, and if $x\in I_{1,2}=[a_2,b_1]$ then the state of the system can be $H_1$ or $H_2$.

Thus, the controlled parameter is the more informative the greater part of the total range of possible values $I_0=[a_1,b_2]$ belongs to the certainty interval $I=[a_1,a_2]\cup[b_1,b_2]=I_0-I_{1,2}$, where $I_{1,2}$ is the compatibility interval. In this case, the information value of the parameter is estimated by the value

$$\varsigma = \left(I_0 - I_{1,2}\right)/I_0 .$$

The measure $\varsigma$ is equal to 0 if the densities $\varphi_1 = (x/H_1)$ and $\varphi_2 = (x/H_2)$ coincide, that is $I_{1,2} = 1$, and is equal to unity if the compatibility interval is $I_{1,2} = 0$. The proposed measure of informativeness is free from the disadvantages of the Kullback criterion and others discussed above, and is easily calculated.

The informativeness of the introduced criterion can be enhanced by value calculation probabilities of the controlled parameter falling into the compatibility interval $I_{1,2}$ for the states $H_1$ and $H_2$ and taking into account their possible differences.

The general disadvantage of the proposed and all other known methods for evaluating the informativeness of parameters is the incomplete accounting of the uncertainty of the initial data, which manifests itself in the inaccurate setting of the range of possible values of the controlled parameter and the inaccuracy of its measurement. Inaccuracies of the first type are eliminated with an increase in the volume of statistical data. The task of accounting for the inaccuracy of measurements of the controlled parameter was not considered. In this regard, the purpose of the work is to develop a methodology for assessing the informativeness of the controlled parameters of the system, taking into account the inaccuracy of the initial data regarding their values.

## Method of controlled parameter information value calculation taking into account the inaccuracy of their measurements

We will solve the problem under the assumption that the inaccuracy of the controlled parameters obtained values is adequately reflected in terms of fuzzy mathematics by a set of membership functions $(L-R)$– such as these values $\mu_{H_1}(x)$ and $\mu_{H_2}(x)$.

We introduce these membership functions of the fuzzy values of the controlled parameter $x$ for the states $H_1$ and $H_2$ of the system: $\mu_{H_1}(x) = < m_1, \alpha_1, \beta_1 >$, that is

$$\mu_{H_1}(x) = \begin{cases} 0, x \le m_1 - \alpha_1; \\ \left(x - (m_1 - \alpha_1)\right)/\alpha_1, m_1 - \alpha_1 < x \le m_1; \\ \left((m_1 + \beta_1) - x\right)/\beta_1, m_1 < x \le m_1 + \beta_1; \\ 0, x > m_1 + \beta_1; \end{cases}$$

$\mu_{H_2}(x) = < m_2, \alpha_2, \beta_2 >$, that is

$$\mu_{H_2}(x) = \begin{cases} 0, x \le m_2 - \alpha_2; \\ \left(x - (m_2 - \alpha_2)\right)/\alpha_2, m_2 - \alpha_2 < x \le m_2; \\ \left((m_2 + \beta_2) - x\right)/\beta_2, m_2 < x \le m_2 + \beta_2; \\ 0, x > m_2 + \beta_2. \end{cases}$$

These relations are illustrated in Fig. 2.

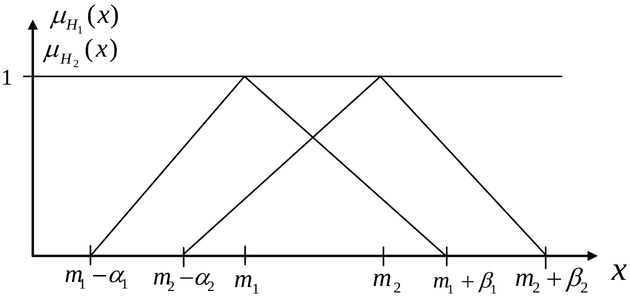


**Fig. 2.** Accessory functions $\mu_{H_1}(x)$ and $\mu_{H_2}(x)$

Let us now take into account the inaccuracy of measuring the values of the controlled parameter $x$. In this regard, we introduce:

$x$ – the true value of the controlled parameter,

$x$ –the measured value of the controlled parameter.

Let $\delta$ be the maximum error of measuring the parameter value.

Then we will assume that the true value of this parameter is a fuzzy number given on the interval $[x-\delta, x+\delta]$ with the membership function $\mu(x) = < x, \delta, \delta >$, that is

$$\mu(x) = \begin{cases} 0, x < x - \delta; \\ \dfrac{x - (x - \delta)}{\delta}, x - \delta < x \le x; \\ \dfrac{(x + \delta) - x}{\delta}, x < x \le x + \delta; \\ 0, x > x + \delta. \end{cases}$$

In order to use the introduced membership functions to calculate the probabilities of the controlled parameter falling into the compatibility interval, we bring these functions to the form adopted in probability theory for the densities of the random variables [7, 8] distribution using known techniques [9, 10].

To do this, calculate the areas under the curves corresponding to $\mu_{H_1}(x), \mu_{H_2}(x)$.

$$S_1 = \int_{m_1-\alpha_1}^{m_1+\beta_1} \mu_{H_1}(x)dx = \int_{m_1-\alpha_1}^{m_1} \frac{x - (x - \alpha_1)}{\alpha_1} dx + \\ + \int_{m_1}^{m_1+\beta_1} \frac{(m_1 + \beta_1) - x}{\beta_1} = S_{1\alpha} + S_{1\beta};$$

$$S_2 = \int_{m_2-\alpha_2}^{m_2+\beta_2} \mu_{H_2}(x)dx = \int_{m_2-\alpha_2}^{m_2} \frac{x - (m_2 - \alpha_2)}{\alpha_2} dx + \\ + \int_{m_2}^{m_2+\beta_2} \frac{(m_2 + \beta_2) - x}{\beta_2} dx = S_{2\alpha} + S_{2\beta}.$$

Let's perform the necessary calculations.

$$S_{1\alpha} = \int_{m_1-\alpha_1}^{m_1} \frac{x-(m_1-\alpha_1)}{\alpha_1}dx = \frac{1}{\alpha_1} \times \frac{x^2}{2}\Bigg|_{m_1-\alpha_1}^{m_1} - \frac{m_1-\alpha_1}{\alpha_1}x\Bigg|_{m_1-\alpha_1}^{m_1} =$$

$$\frac{1}{\alpha_1}\left(\frac{m_1}{2} - \frac{(m_1-\alpha_1)^2}{2}\right) - \frac{m_1-\alpha_1}{\alpha_1}(m_1 - m_1 + \alpha_1) =$$

$$= \frac{1}{2\alpha_1}\left(\cancel{m_1}^2 - \cancel{m_1}^2 + 2m_1\alpha_1 - \alpha_1^2\right) - \frac{(m_1-\alpha_1)\alpha_1}{\alpha_1} = m_1 - \frac{\alpha_1}{2} - m_1 + \alpha_1 = \frac{\alpha_1}{2};$$

$$S_{1\beta} = \int_{m_1}^{m_1+\beta_1} \frac{(m_1+\beta_1)-x}{\beta_1}dx = \frac{m_1+\beta_1}{\beta_1}x\Bigg|_{m_1}^{m_1+\beta_1} - \frac{1}{\beta_1}\cdot\frac{x^2}{2}\Bigg|_{m_1}^{m_1+\beta_1}$$

$$= \frac{(m_1+\beta_1)}{\beta_1}(\cancel{m_1} + \beta_1 - \cancel{m_1}) - \frac{1}{2\beta_1}\left[(m_1+\beta_1)^2 - m_1^2\right] = m_1 + \beta_1 - \frac{1}{2}\beta_1 \times$$

$$\left(\cancel{m_1}^2 + 2m_1\beta_1 + \beta_1^2 - \cancel{m_1}^2\right) = m_1 + \beta_1 - m_1 - \frac{\beta_1}{2} = \frac{\beta_1}{2}.$$

Then

$$S_1 = \frac{\alpha_1}{2} + \frac{\beta_1}{2} = \frac{\alpha_1+\beta_1}{2}. \qquad (14)$$

The result is quite predictable, since the corresponding figure $\mu_{H_1}(x)$ is a triangle with a height equal to one and a base $\alpha_1 + \beta_1$. From here the result of integration (5) can be written:

$$S_2 = \frac{\alpha_2+\beta_2}{2}.$$

In order to simplify the recording of the following relations, we introduce:

$a = (m_2 - \alpha_2)$ – the left boundary of the controlled parameter true values compatibility interval;

$b = (m_1 + \beta_1)$ — the right boundary of controlled parameter true values compatibility interval;

$P_1 = (x - \delta)$ — the left boundary of the controlled parameter true values interval;

$P_2 = (x + \delta)$ — the right boundary of controlled parameter true values interval.

Now we formulate the following task: calculate the probability of a controlled parameter $[a,b]$ falling into the compatibility interval for states $H_1$ and $H_2$ that is

$$P(x \in I_{1,2}) = P(x \in [a,b]). \qquad (15)$$

Calculated ratios for probabilities (15) determination are written differently depending on the position of the observed parameter value in relation to the compatibility interval.

Possible variants of controlled parameter true values interval position within the compatibility interval are shown in Fig. 3.1-3.3.

In this case, the compatibility interval $[a,b]$ is divided into three subintervals: left boundary – $[a, a+\delta]$, central – $[a+\delta, b-\delta]$, right boundary – $[b-\delta, b]$.

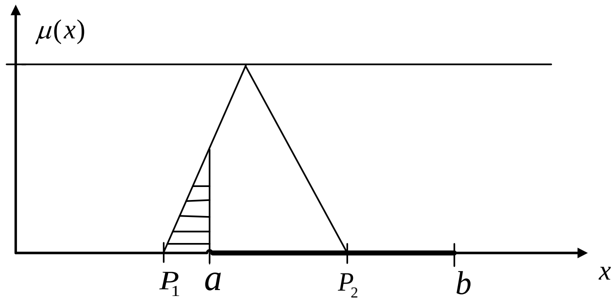


**Fig. 3.1** Option: $a < x < a + \delta$

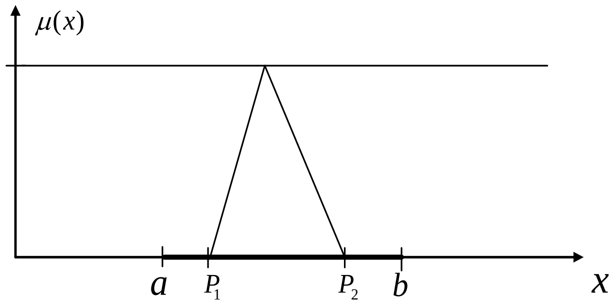


**Fig. 3.2** Option: $a + \delta \le x \le b - \delta$

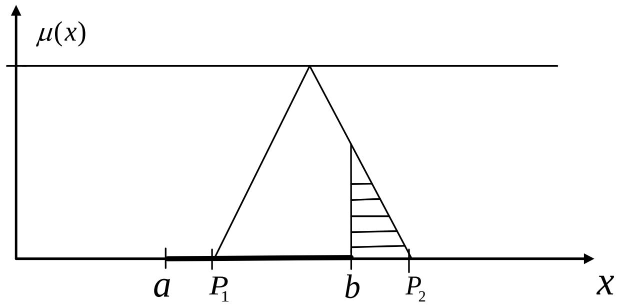


**Fig. 3.3** Option: $b - \delta < x < b$

The corresponding probabilities of hitting the values of the controlled parameter for the states $H_1$ and $H_2$ are calculated by the formulas:

$$P_{H_2}{}^{(1)} = \frac{1}{S_2}\int_a^{a+\delta}\mu_{H_2}(x)dx = \frac{1}{S_2}\int_a^{a+\delta}\frac{x-(m_2-\alpha_2)}{\alpha_2}dx = \frac{1}{S_2\alpha_2}\left[\frac{x^2}{2}\bigg|_a^{a+\delta} - (m_2-\alpha_2)x\bigg|_a^{a+\delta}\right] =$$
$$= \frac{1}{S_2\alpha_2}\left[\frac{(a+\delta)^2}{2} - \frac{a^2}{2}\bigg| \; -(m_2-\alpha_2)(a+\delta-a)\right] = \frac{1}{2S_2\alpha_2}\left[\left(2a\delta+\delta^2\right)-\delta(m_2-\alpha_2)\right]; \quad (16)$$

$$P_{H_1}{}^{(1)} = \frac{1}{S_1}\int_a^{a+\delta}\mu_{H_1}(x)dx = \frac{1}{S_1}\int_a^{a+\delta}\frac{(m_1+\beta_1)-x}{\beta_1}dx = \frac{1}{S_1\beta_1}\left[x(m_1+\beta_1)\bigg|_a^{a+\delta} - \frac{x^2}{2}\bigg|_a^{a+\delta}\right] = \frac{1}{S_1\beta_1}\times$$
$$\times\left[\left((a+\delta)-a\right)(m_1+\beta_1) - \frac{(a+\delta)^2-a^2}{2}\right] = \frac{1}{2S_1\beta_1}\left[\left(2\delta(m_1+\beta_1)\right)-\left(2a\delta+a^2\right)\right] = \frac{1}{2S_1\beta_1}\left[2\delta(m_1+\beta_1-a)-a^2\right]; \quad (17)$$

$$P_{H_1}{}^{(2)} = \frac{1}{S_1}\int_{a+\delta}^{b-\delta}\mu_{H_1}(x)dx = \frac{1}{S_1}\int_{a+\delta}^{m_1}\frac{x-(m_1-\alpha_1)}{\alpha_1}dx + \frac{1}{S_1}\int_{m_1}^{b-\delta}\frac{(m_1+\beta_1)-x}{\beta_1}dx = \frac{1}{S_1\alpha_1}\times$$
$$\times\left[\frac{x^2}{2}\bigg|_{a+\delta}^{m_1} - (m_1-\alpha_1)x\bigg|_{a+\delta}^{m_1}\right] + \frac{1}{S_1\beta_1}\left[(m_1+\beta_1)\bigg|_{m_1}^{b-\delta} - \frac{x^2}{2}\bigg|_{m_1}^{b-\delta}\right] = \frac{1}{S_1\alpha_1}\left[\frac{m_1^2}{2} - \frac{(a+\delta)^2}{2} - (m_1-\alpha_1)(m_1-a-\delta)\right] + \quad (18)$$
$$+\frac{1}{S_1\beta_1}\left[\left[(m_1+\beta_1)(b-\delta)-(m_1+\beta_1)m_1\right] - \left[\frac{(b-\delta)^2}{2} - \frac{m_1^2}{2}\right]\right];$$

$$P_{H_2}{}^{(2)} = \frac{1}{S_2}\int_{a+\delta}^{b-\delta}\mu_{H_2}(x)dx = \frac{1}{S_2}\int_{a+\delta}^{m_2}\frac{x-(m_2-\alpha_2)}{\alpha_2}dx + \frac{1}{S_2}\int_{m_2}^{b-\delta}\frac{(m_2+\beta_2)-x}{\beta_2}dx =$$
$$= \frac{1}{S_2\alpha_2}\left[\frac{x^2}{2}\bigg|_{a+\delta}^{m_2} - (m_2-\alpha_2)x\bigg|_{a+\delta}^{m_2}\right] + \frac{1}{S_2\beta_2}\left[(m_2+\beta_2)x\bigg|_{m_2}^{b-\delta} - \frac{x^2}{2}\bigg|_{m_2}^{b-\delta}\right] =$$
$$= \frac{1}{S_2\alpha_2}\left[\frac{m_2^2}{2} - \frac{(a+\delta)^2}{2} - (m_2-\alpha_2)m_2 + (m_2-\alpha_2)(a+\delta)\right] + \quad (19)$$
$$+\frac{1}{S_2\beta_2}\left[(m_2+\beta_2)(b-\delta)-(m_2+\beta_2)m_2 - \frac{(b-\delta)^2}{2} + \frac{m_2^2}{2}\right];$$

$$P_{H_1}{}^{(3)} = \frac{1}{S_1}\int_{b-\delta}^{b}\mu_{H_1}(x)dx = \frac{1}{S_1}\int_{b-\delta}^{b}\frac{(m_1+\beta_1)-x}{\beta_1}dx = \frac{1}{S_1\beta_1}\left[(m_1+\beta_1)x\bigg|_{b-\delta}^{b} - \frac{x^2}{2}\bigg|_{b-\delta}^{b}\right] =$$
$$= \frac{1}{S_1\beta_1}\left[(m_1+\beta_1)(b-b+\delta) - \frac{b^2}{2} + \frac{(b-\delta)^2}{2}\right]; \quad (20)$$

$$P_{H_2}{}^{(3)} = \frac{1}{S_2}\int_{b-\delta}^{b}\mu_{H_2}(x)dx = \frac{1}{S_2}\int_{b-\delta}^{b}\frac{(m_2+\beta_2)-x}{\beta_2}dx = \frac{1}{S_2\beta_2}\left[(m_2+\beta_2)x\bigg|_{b-\delta}^{b} - \frac{x^2}{2}\bigg|_{b-\delta}^{b}\right] =$$
$$= \frac{1}{S_2\beta_2}\left[(m_2+\beta_2)(b-b+\delta) - \frac{b^2}{2} + \frac{(b-\delta)^2}{2}\right] = \frac{1}{S_2\beta_2}\left[(m_2+\beta_2)\delta - \frac{b^2}{2} + \frac{(b-\delta)^2}{2}\right]. \quad (21)$$

## Discussion of the results obtained

It is shown that the results of calculations using formulas (16) – (21) contain important information about the state of the system in a critical situation when the measured value of the controlled parameter is within the compatibility interval. Comparison of the results of probabilities calculations falling into the left boundary subinterval (ratios (16) - (17)) for different states of the system, into the central subinterval (ratios (18) - (19)) and the right boundary subinterval (ratios (20) - (21)) are all the more informative for these states, the more significant the difference in the numerical values of the corresponding probabilities. The practical significance of the proposed technique is especially important if the compatibility interval for competing diagnoses is a significant part of the total range of possible values of the controlled parameter.

## Conclusions

The possibility of increasing the accuracy of assessing system controlled parameters information importance in the task of its state identification is considered.

It is shown that taking into account and analyzing the probabilities of the values of the controlled parameters falling into the range of their possible values common for different states provides significant additional information about the informational value of these parameters.


REFERENCES

1. Afifi ,A, and Eisen, S. (1982), *Statistical analysis*, MIR, Moscow, 488 p.
2. Duran, B. and Odell, P. (1977), *Cluster analysis*, Statistics, Moscow, 128 p.
3. Mandel, I D. (1988), *Cluster analysis*, Finance and statistics, Moscow, 176 p.
4. Cornwell, B. (2015), "Linkage Criteria for Agglomerative Hierarchical Clustering", *Social Sequence Analyse*, 2015, pp.270–274, doi: https://doi.org/10.1017/CBO9781316212530.011
5. Kullback, S. and Leibler, R.A. (1951), "On information and Sufficiency", Ann. Math. Statist, Vol. 22, No. 1, pp. 79–86, available at: https://www.jstor.org/stable/2236703
6. Semenov, S., Sira, O., Gavrylenko, S. and Kuchuk, N. (2019), "Identification of the state of an object under conditions of fuzzy input data", *Eastern-European Journal of Enterprise Technologies*, Vol 1, No 4 (97), pp. 22-30, doi: https://doi.org/10.15587/1729-4061.2019.157085
7. Feller, W. (1970), An introduction to probability theory and applications, J. Wiley, 527 p., available at: https://bitcoinwords.github.io/assets/papers/an-introduction-to-probability-theory-and-its-applications.pdf
8. Cramer, H. (1975), Mathematical methods of statistics, Stockholm, 647 p.
9. Raskin, L.G. and Pustovoitov, P.E. (2002), "Solution of a multi-product problem of inventory management by a probabilistic criterion", *Bulletin of NTU KhPI, Series System analysis, management, information technology*, No. 13, pp. 49–53.
10. Raskin, L., Sira, O., & Ivanchykhin, Y. (2017), "Models and methods of regression analysis under conditions of fuzzy initial data", *Eastern-European Journal of Enterprise Technologies*, Vol, 4(4 (88)), pp. 12–19, doi: https://doi.org/10.15587/1729-4061.2017.107536



Received (Надійшла) 14.12.2022
Accepted for publication (Прийнята до друку) 08.02.2023



ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

**Раскін Лев Григорович**– доктор технічних наук, професор, професор кафедри мультимедійних інформаційних технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
**Lev Raskin** – Doctor of Technical Sciences, Professor, Professor in the Department of Multimedia Information Technology and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: topology@ukr.net; ORCID: http://orcid.org/0000-0002-9015-4016.

**Сухомлин Лариса Вадимівна** – кандидат технічних наук, доцент, доцент кафедри менеджменту, Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна;
**Larysa Sukhomlyn** – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Management, Kremenchuk Mikhail Ostrogradskiy National University, Kremenchuk, Ukraine;
e-mail: lar.sukhomlyn@gmail.com; ORCID: https://orcid.org/0000-0001-9511-5932.

**Соколов Дмитро Дмитрович** – аспірант кафедри мультимедійних інформаційних технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
**Dmytro Sokolov** – postgraduate student in the Department of Multimedia Information Technology and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: sokolovddd@gmail.com; ORCID: https://orcid.org/0000-0002-4558-9598.

**Власенко Віталій Владиславович** – магістр кафедри мультимедійних інформаційних технологій і систем, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна;
**Vitalii Vlasenko** – master of the Department of Multimedia Information Technologies and Systems, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: vitalik.vlasenko.000@gmail.com; ORCID: https://orcid.org/0000-0001-5427-0223.


### Оцінка інформаційної важливості контрольованих параметрів системи з урахуванням неточності вихідних даних

Л. Г. Раскін, Л. В. Сухомлин, Д. Д. Соколов, В. В. Власенко

**Анотація.** У роботі розглянуто технологію оцінки інформаційної цінності контрольованих параметрів системи в задачі ідентифікації її стану. **Мета дослідження** – удосконалення стандартної методики оцінки інформаційної цінності контрольованих параметрів. Запропонований метод заснований на аналізі ймовірностей потрапляння значення контрольованого параметра в підінтервали інтервалу можливих значень різних станів системи. При попаданні значення контрольованого параметра в лівий або правий граничні підінтервали інтервалу сумісності для будь-якого стану об'єкта висновок про його стан робиться з урахуванням можливих у цьому випадку помилок першого або другого роду. При попаданні значення контрольованого параметра центральний підінтервал корисна інформація з'являється, якщо відповідні ймовірності для станів $H_1$ nf $H_2$ істотно відрізняються. Таким чином, показано, що врахування ймовірностей попадання нечітких значень контрольованого параметра в інтервал сумісності для різних станів об'єкта суттєво підвищує його інформаційну цінність.

**Ключові слова:** завдання ідентифікації станів системи; інформаційна цінність контрольованих параметрів; врахування ймовірностей їх потрапляння до інтервалів можливих значень..