Uncertainties for symmetric and asymmetric distribution models

Vitaly G. Polosin*

Medical Cybernetics and Computer Science Department, Penza State University, 40, Krasnay Str., Penza, 440026, Russia

Abstract. The model of the symmetric distribution shows the relationship between both the intervals of entropy uncertainty and the peculiarities of the formation of intervals of parametric uncertainty for symmetric and non-symmetric distributions. In particular, the proportionality of the entropies of symmetric and non-symmetric distributions is shown. The peculiarity of the formation of entropy features of symmetric and asymmetric distribution is shown. The entropy coefficient of asymmetric distributions is proposed as an independent entropy feature of the shape of asymmetric distributions. The peculiarity of applying the entropy coefficient of a symmetric distribution to the description of the properties of an asymmetric distribution is discussed. In this paper the peculiarity of applying the entropy coefficient of a symmetric distribution to the description of the properties of an asymmetric distribution also is discussed. In particular, it is shown that the use of the entropy feature of a symmetric distribution can be considered as a replacement for an asymmetric distribution near the estimates of the center of random variables by means of its symmetric model intervals of parametric and entropy uncertainty. Such a replacement is possible if the intervals of the models are equal.

1 Introduction

Mathematical models as a method for studying phenomena and objects of the real world are used in various fields of activity of modern society. The use of statistical models in all areas of knowledge it is characteristic of the modern stage of development of natural and technical sciences as due to the intensive development of information technologies and computing facilities. The task of constructing a model in any field of research is to identify patterns of govern real processes in physics, technical, economic, medical, environmental and other spheres of human activity. Since the found patterns contains both theoretical and practical value, they are possible to use in planning, forecasting, control and management problems. The random nature of the change in the state of real systems necessitates the use of probabilistic patterns for constructions of decisions-making models on its control. The state and behavior of complex systems determine the parameters of probabilistic models of stochastic processes.

* Corresponding author: polosin-vitalij@yandex.ru

Modern methods of model selection are based both on probabilistic and informational analysis of the sample uncertainty of the observed output quantity. These methods have already proven themselves in different fields of scientific and applied research as such estimated the measurement error [1], measuring non-electrical quantities [2], automation and telemechanic systems [3], communication systems [4], organizing monitoring and control processes of dynamic systems [5], control of complex systems [6], studies of the activity of radiation sources [7] and others. For this reason, the study of uncertainty intervals of symmetric and non-symmetric models of distributions of random variables is relevant for the processing and analysis of random processes. The author of this paper considers the features of the construction of uncertainty coefficients for symmetric and non-symmetric distributions.

2 Measures of the space of elementary event

In problems of probabilistic modeling, the space of elementary events of Ω to assess the uncertainty of random variables is used, that represents the set of all mutually exclusive outcomes of a random experiment. If the set of elementary events is of course, then discrete space is used to describe events. As a measure of the space of elementary event, the measure of probability is used, that is given by the probability distribution in the form of a real non-negative function $P(\Omega)$ on the algebra of the space of elementary event [8]. The probability measure $P(\Omega)$ is finitely additive and normalized by one, that is equal to one over the entire space.

A sample of an n-dimensional random variable with independent components is given by a vector of dimension n that is determines the position of the n-dimensional vector of a sample of a random variable in the space of elementary event. Since the events are independent in the space of elementary event, the distance ρ(X, Y) between the samples of values of (x_1, x_2, … x_n) and (y_1, y_2, … y_n) of the X and Y random variables is the Euclidean measure that is equal as

$$\rho(X,Y) = \|x-y\| = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}. \quad (1)$$

If the y_i values for the vector of the Y random variable of formula (1) are taken equal to zero, then the ratio of the squared distance to the dimension of the vector n is equal to the second sample initial moment of the X random variable that is given as

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2. \quad (2)$$

The product of the second central moment and the dimension of the sample of the random variable, it is equal to the square of the radius vector of the position of the n-dimensional sample of the random variable in the Euclidean space of elementary events. The square root of the second initial moment is equal to the mean square value of the random variable.

It is convenient to characterize the center of the probability distribution using the mathematical expectation equal to the first initial moment m1. If the y_i values of the Y vector are equal to the mathematical expectation m1 of the X random variable, then the ratio of the Euclidean measure (1) to the n dimension of the X vector corresponds to the second central

$$\mu_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-m_1)^2. \quad (3)$$

The product of the second central moment by the dimension of the vector of a random variable is the distance between the estimate of the center of sample and its true position in the Euclidean space of elementary events. The square root of the second central moment is equal to the biased estimate of the sample variance of the random variable.

Thus, the second central and initial moments of samples of random variables should be considered as Euclidean measures in the space of elementary events, which are normalized with respect to the dimension of the sample.

The constructing of Euclidean measures is based on the second central moment and on the second initial moment of distributions, it made is possible to consider the sample of results as an n-dimensional Euclidean space of elementary events. The application possibility of the selective moments for research of distributions properties is often discussed that there are in papers [8, 9].

3 A measure of distribution density information

The entropy of information is used by C. Shannon as a measure for measuring the amount of information in an ensemble of discrete message. If the source of information produces a random sequence of x_i independent messages with probability $P(x_i)$, then the entropy of the messages X source is determined by the mathematical expectation of the logarithm of the messages probability that is given as [10, 11]

$$H(X) = M(-\ln p\,(x)). \tag{4}$$

When a measurement will be making, the properties of an object are characterized by continuously distributed values with a density that is $f\,(x)$. The $p(x \pm \Delta x)$ probability of the appearance of any value in the $(\pm\Delta x)$ interval is equal to the product of the distribution density function in this interval by the value of the interval. Then the entropy of the discrete signal is determined through the sums of the logarithms

$$H(X) = M(-\ln f\,(x)) + M(-\ln 2\,\Delta x). \tag{5}$$

The first term is the differential entropy of the distribution density function of $f(x)$, it is used as a finite characteristic of the entropy of a continuous source [14]. The second term in expression (5) reflects the entropy of the signal sampling interval and does not depend on the distribution of the random variable. For a continuous source the second term tends to infinity [12]. During measurement, we receive information about the object. The I information received quantities is equal to the difference in entropies before and after the measurement [1]. It is given by

$$I = H(X) - H(X/x_p). \tag{6}$$

Where $H(x/x_p)$ is the conditional entropy, which is determined under the condition that the x_p measurement result is obtained.

From expression (6) it follows that during the measurement there is the change in the quantities of information, which is equal to the difference in the differential terms of the signal entropies.

4 Uncertainty intervals for symmetric distribution density of a random variable

It was proposed in the paper of P.V. Novitsky [1] The uncertainty model for symmetric distributions use as a model for the independent description of the measurement process. In this model, the condition was applied that a meter reading of x_p was obtained with a uniform distribution of uncertainty over the interval $2H_s$. Then if the x_p meter reading was obtained then the conditional entropy of measurement is equal to the logarithmic measure of the entropy interval of uncertainty that is $2H_s$. It was given as

$$H(X/x_p) = -\int_{x_p-\Delta_{Hs}}^{x_p+\Delta_{Hs}} \frac{1}{2\Delta_{Hs}} \ln \frac{1}{2\Delta_{Hs}}\, dx = \ln(2\Delta_{Hs}) \tag{7}$$

The entropy measurement error is taken to be equal to half of the uncertainty interval 2, which is expressed through the conditional entropy of change, provided that the x_p reading is obtained. This is given by

$$H(X/x_p) = -\int_{x_p-\Delta_{Hs}}^{x_p+\Delta_{Hs}} \frac{1}{2\Delta_{Hs}} \ln \frac{1}{2\Delta_{Hs}}\, dx = \ln(2\Delta_{Hs}) \tag{8}$$

In control and management systems of an object, the minimization of the difference in the mismatch between the output parameter and the measure by constructing negative feedback is ensured. As the measure of the uncertainty of the control system, the paper [5] also used half of the entropy uncertainty interval.

At figure 1 uncertainty models for a symmetric distribution of a random variable are shown. These models illustrate the normal Gaussian distribution of the number 1 and the uniform distribution of the number 2. The normal distribution was constructed from the condition that the standard deviation σ of the random variable is equal to one. The uniform distribution was constructed from the condition that the differential entropies of normal and uniform distributions are the same. The entropy of the normal distribution is determined by the expression.

$$H_N(X/x_p) = ln\left(\sigma_N\sqrt{2\pi e}\right). \quad (9)$$

If the entropy measures are equal, then length of the uniform distribution sets an independent finite interval for the uncertainty of a random variable for the Gaussian normal distribution. Similarly, a relationship between a uniform symmetric distribution and any other distribution with symmetric density is build. For this reason, it is convenient to use the $2\Delta_{Hs}$ interval as an informational estimate of the uncertainty interval of symmetrically distributed random variables relative to their center

In the space of elementary events, it is generally accepted to use the standard deviation as the Euclidean measure of uncertainty. Since the ratio of information and probabilistic uncertainties characterizes the form of symmetric distributions, the entropy coefficient is used to analyze symmetric distributions. In order to determining the entropy coefficient, the equation is given:

$$K_{Hs} = \frac{\Delta_{Hs}}{\sigma}. \quad (10)$$

It is shown in Novitsky's paper that the entropy coefficient (4) is an effective estimate of symmetric distributions. If both the entropy coefficient and kurtosis are used as estimates of the shape, then this allows one to identify the shape of the symmetric distribution. In fact, application of expression (15) to non-symmetric distributions of random variables is reduced to replacing the non-symmetric distribution by its symmetric model. In order to illustrate the replacement, Figure1 shows an asymmetric gamma distribution with shape parameter and scale parameter that are equal to 4.4 and 1.9, respectively.

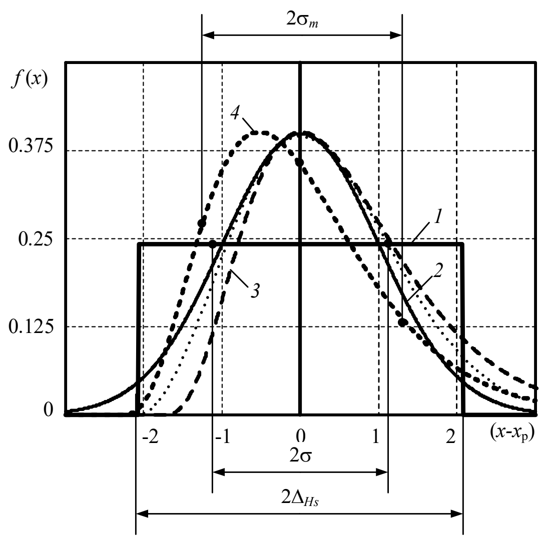

Fig. 1. Uncertainty model by symmetric distribution of random variable.

Figure 1 also demonstrates curves 3 and 4, which illustrate the position of the non-symmetric distribution of the relative position of the center of the uniform distribution when respectively the mode or the mean value as the center estimates are applied. In the case of a skewed distribution, the skewed distribution is replaced by a symmetric model with respect to the result obtained.

5 Uncertainty intervals at non-symmetric distribution density of a random variable

For distributions with the non-symmetric unbiased distribution density, a distinctive feature is that the probability density of the distribution is zero for negative values of the random variable. It is obvious that the relationship between the intervals of probabilistic and information uncertainty will change if the distribution values are specified asymmetrically and are located only on one semi axis. According to the theory of probability, the mathematical expectation of the probability logarithm of a random variable can be written in the form of a functional.

$$H(X) = -\int_{-\infty}^{\infty} f(x)\, ln\, f\,(x) dx. \tag{11}$$

In the negative range of values, the limiting value for the integral expression (11) is equal to zero. Indeed, it is given as $\lim_{x\to 0}(x \ln x) = \lim_{x\to 0}(\ln x^x) = \ln 0^0 = \ln 1 = 0$.

Then the functional for determining the entropy of a random variable with an non-symmetric distribution density is given as

$$H_n(X) = -\int_0^{\infty} f_n(x)\, ln\, f_n\,(x) dx. \tag{12}$$

Where $f_n(x)$ is non-symmetric distribution density. Examples of the position of distributions of a random variable with asymmetric and unbiased distribution densities it is shown at Figure 2. The following designations are given herein. The number 1 is a uniform distribution.

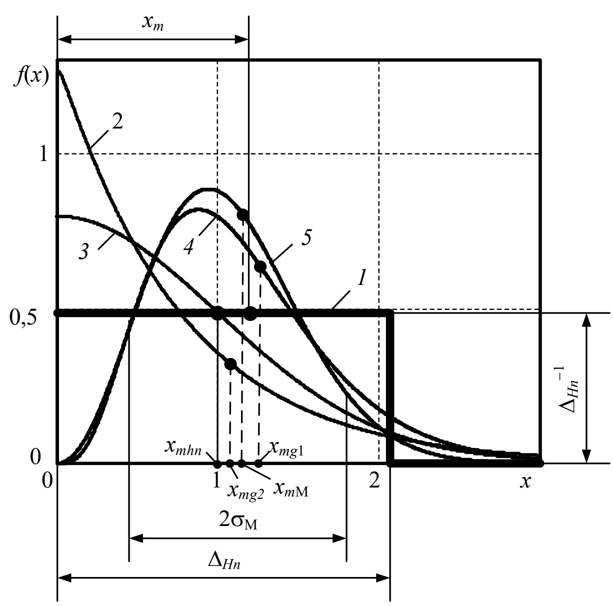

Fig. 2. Uncertainty model for non-symmetric distribution of a random variable.

The number 2 is the gamma distribution with the shape and span parameter that are equal to 1.01 and 1.32 respectively. The number 3 is a semi-normal distribution that was reflected relative to the distribution center. The number 4 is the gamma of the distribution with the shape and span parameter that is equal to 4.4 and 3.87 respectively. The number 5 is the Maxwell distribution.

It is convenient to characterize the uncertainty of the distribution of random variables using an interval of uniform distribution. Since asymmetric distributions are used for physical quantities that it is take only positive values, then it is possible to use for set the uncertainty of non-symmetric distributions a uniform distribution that is also located non-symmetrically relative to the origin in the range from 0 to Δ_{Hn}. It is to determine the uncertainty interval of the uniform distribution. This is required that the entropies of the H_u uniform distribution and the H_n simulated asymmetric distribution will be equal. Then the interval of uniform distribution can be considered as the entropy interval of the uncertainty of a random variable.

Based on normalization condition, this was obtained that the uniform distribution density is inversely proportional to the uniform distribution interval. If a random variable is distributed in the range from 0 to H_n, then the expression for the distribution density is given as

$$f_1(x) = {\Delta_{Hn}}^{-1} \tag{13}$$

Then the entropy of uniformly distributed of the random variable, is equal to the logarithm of the range of its distribution. The formula for calculating the entropy uncertainty interval for an asymmetric distribution with a known entropy is given as

$$\Delta_{Hn} = e^{H_n(X)} \tag{14}$$

From a comparison of the obtained expression (14) and expression (8), it follows that for the same values of the entropies of symmetric and non-symmetric distributions, the interval of entropy uncertainty H_n of the non-symmetric distribution of a random variable is twice as large as the interval of entropy uncertainty H_s of symmetric random variable.

6 Another section of your paper

The relationship between the uncertainty intervals of symmetric and non-symmetric distributions can be illustrated on the model of symmetric distribution, which is obtained by reflecting the non-symmetric unbiased distribution to the region of negative values.

Let the $f_n(x)$ density function of the non-symmetric distribution be given on the interval of positive values. If $f(y)$ is the density function of the model of the symmetric distribution of the Y random variable, that it is obtained by mapping the $f_n(x)$ density function the nonsymmetric distribution of the X random variable, which is specified only on the positive semiaxis, then the expression for calculating the $f(y)$ density function will be given as

$$f(y,\alpha,\tau,\lambda) = \begin{cases} 0.5 \cdot f_n(y,\alpha,\tau,\lambda) & by \quad y \geq 0, \\ 0.5 \cdot f_n(|-y|,\alpha,\tau,\lambda) & by \quad y < 0. \end{cases} \tag{15}$$

Where α and τ are the shape parameters of the non-symmetric distribution, λ is the scale parameter of distributions.

An example of constructing a model of a symmetric distribution is illustrated by curves with numbers 2, 3, 4 and 5 in Figure. 3. Models of symmetric distributions were obtained by the following actions. First, the symmetric transfer of the density of non-symmetric distributions with respect to the origin of coordinates was performed, then the densities of symmetric model was normalized over area at unit. Both Figure 2 and Figure 3 contain the same designations for asymmetric distributions and for their symmetric models. If the uncertainty intervals of asymmetric distributions is determine, then it is given the property of proportionality of the entropy of the non-symmetric unbiased distribution with respect to the entropy of the model of the symmetric distribution. These properties by reflecting the asymmetric distribution about the origin were obtained.

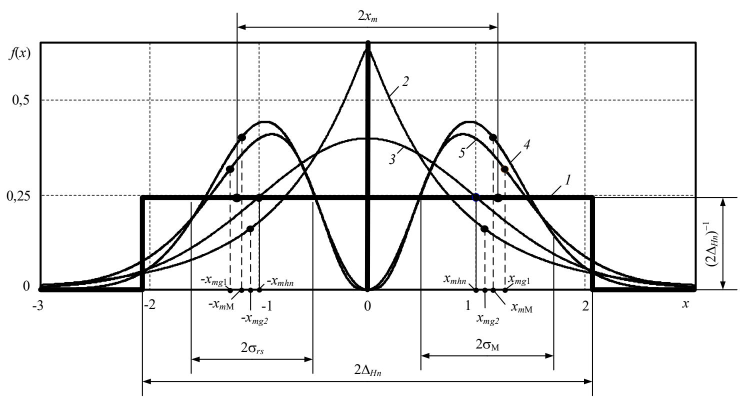

Fig. 2. Densities of models of symmetric distributions of a random variable *Y*.

If let it be used the relationship between distribution densities (15), then it is made to possible the relationship building between the entropies of symmetric and non-symmetric distributions. For these purposes, it is necessary to perform the following actions. First, the density of the symmetric model of the normalized distribution (15) into the expression for functional (9) was used. Then, the substitution of (Z=–Y) for negative values of the Y random variable is apply. There is a negative sign when replacing differentials (dz=–dy), which was taken into account when changed the integration limits in accordance with the used replacement.

As a result of the performed actions, it is obtained an expression for determining the HS entropy of the symmetric distribution model at a known non-symmetric distribution density of $f_n(Y, \alpha, \tau, \lambda)$, that is specified on the positive semiaxis. This expression is given by

$$H_S = -\int_0^\infty \frac{f_n(z,\alpha,\beta,\lambda)}{2} \ln \frac{f_n(z,\alpha,\beta,\lambda)}{2} dz - \int_0^\infty \frac{f_n(Y,\alpha,\beta,\lambda)}{2} \ln \frac{f_n(Y,\alpha,\beta,\lambda)}{2} dy. \quad (16)$$

On the right side of expression (16), random variables of Y and Z are distributed identically in accordance with the law of non-symmetric distribution with the same shape and scale parameters. Then, for the entropy of a symmetric distribution, an expression is valid that is given as

$$H_S = -\int_0^\infty f_n(y,\alpha,\beta,\lambda) \ln \frac{f_n(y,\alpha,\beta,\lambda)}{2} dy. \quad (17)$$

Since the logarithm of the ratio on the right side of the written expression (17) is equal to the difference of the logarithms, the integral expression on the right side of expression (17) can be written as the difference of integrals.

Then one of the terms on the right-hand side of expression (17) corresponds to the right-hand side of the entropy of the asymmetric unbiased distribution that is given by the expression (12). The integral of the density function of the other term is equal to unity, since the integration is carried out over an asymmetric unbiased distribution density in the range from zero to infinity.

Thus, the validity of linear proportionality for the entropy of the model of symmetric and for the entropy of asymmetric distributions, which is given by

$$H_S(Y) = H_n(X) + \ln 2. \quad (18)$$

According to the obtained expression (18), if the model of the symmetric distribution obtained as a result of symmetric transfer relative to the origin of the density of the non-symmetric distribution that was only determined on the positive axis, then the entropy of the model of the normalized symmetric distribution is equal to the entropy of the non-symmetric unbiased distribution of the number 2 added with the logarithm.

If the operation of potentiation of expression (18) will be carry out and then the potentials of the entropies through the intervals of uncertainty of these entropies in accordance with expressions (8) and (14) is replace, then is get the proportionality of the intervals. It is given

$$\Delta_{Hs} = 2\Delta_{Hn}. \tag{19}$$

From the obtained relation (19) it follows that the uncertainty intervals of the symmetric distribution model are twice as large as the uncertainty interval of the non-symmetric unbiased distribution, which is specified only on the positive semiaxis.

7 Entropy coefficient of unshifted asymmetric distribution

The relationship between the entropy properties of non-symmetric distribution and its model of symmetric distribution is used by the author these papers to construct the entropy coefficients of non-symmetric distributions. To clarify the materials, Figure 3 shows an illustration of the position of the intervals of parametric and entropic uncertainties. The parametric uncertainty interval of the symmetric distribution is specified as twice the standard deviation for the *Y* symmetric random variable. The parametric uncertainty of the symmetric distribution characterizes the scatter of the *Y* random variable relative to the position of the center of the symmetric distribution, which is equal to zero.

For the *S* standard scatter of the symmetric distribution of the *Y* random variable, the following formula is valid

$$\sigma_S = \sqrt{\int_{-\infty}^{\infty} y^2 \cdot \frac{1}{2} f_n(|y|, \alpha, \beta, \lambda) dy}. \tag{20}$$

The second initial moment of the non-symmetrically distributed of *X* random variable is given as

$$m_2 = \int_0^{\infty} x^2 f_n(x, \alpha, \beta, \lambda) dx. \tag{21}$$

On the right-hand side of expression (20), integration is performed over a function that does not depend on the sign of the *Y* random variable. Therefore, it is obtaining the result of integration in expression (20), it is sufficient to double the value of the integral that was obtained only over the region of positive values of the y-axis. Since the integrals in expression (20) and (21) are equal, the equality is true that is given by

$$\sigma_S = \sqrt{m_2}. \tag{22}$$

Thus, the standard scatter for model of the symmetrical distribution of the random variable *Y*, that obtained by reflection relative to the origin and subsequent normalization of the distribution of the parameter *x*, is equal to the square root of the second initial moment calculated for the original asymmetric distribution of the random variable *X*. For the model of symmetric distribution both the $2H_n$ interval of entropic uncertainty, and the 2*S* interval of parametric uncertainty characterize the uncertainty of the distribution of the *Y* value relative to the center of symmetry. These intervals are located at the figure 3 symmetrically about the origin. For this reason, for the model of the distribution of the *Y* random variable that is symmetric with respect to the origin, the author uses the entropy coefficient of the symmetric distribution, that is equal to the ratio of half of the entropy uncertainty interval of the model of symmetric distribution to its standard scatter. The formula for calculating the entropy coefficient of a symmetric distribution model is given

$$K_{HS} = \frac{\Delta_{Hn}}{\sigma_S}. \tag{23}$$

If the standard deviation in expression (23) the by the square root of the second initial moment according to (22) will be replace, then the expression for the entropy coefficient of the non-symmetric distribution of the random variable is obtained that is given

$$K_{HS} = \frac{\Delta_{Hn}}{\sigma_S}. \quad (24)$$

The right side of expression (18) contains only the parameters of asymmetric distributions that is illustrated in Figure 2. The square root of the second starting moment is equal to the mean square value of the random variable. In Figure 2 it is given for the following distributions. The positions of the mean square values of the uniform and semi-normal distributions correspond to the coordinates of the points of x_m and x_{mhn}. The coordinates of the points of x_{mg2} and x_{mg1} illustrate the change in the positions of the root-mean-square values when the shape parameter of the gamma distribution changes, which are equal to 1.01 and 4.4, respectively. The position of the point with the coordinate x_{mhm} show the root-mean-square values of the well-known Maxwell distribution.

Figure 2 and Figure 3 are showed that all distributions are plotted with equal entropies. Intervals of information uncertainty of these distributions illustrate the boundaries of a uniform distribution. The ratio of entropy and parametric uncertainties depends only on the shape of the distributions.

Thus, the entropy coefficient of asymmetric unshifted distributions is a dimensionless normalized indicator of the shape of a non-symmetric distribution, which is equal to the ratio of the H_n uncertainty interval of the asymmetric distribution of the random variable X to the square root taken from the m_2 second initial moment of the distribution.

If the entropy coefficient of an asymmetric distribution together with both asymmetry and kurtosis is used, creation of new tools for systematizing and analyzing of distributions will be possible.

References

1. P.V. Novitsky, A.I. Zograf, *Error estimation of measurement results* (Leningrad, Energoatomizdat, 1985).
2. A.N. Turchin, P.V. Novitsky, E.S. Lenshina, *Electrical measurements are not electrical quantities* (Leningrad, Energy, 1975).
3. A.V. Fremke, *TV measurement* (Moscow, Higher School, 1968).
4. G.V. Gorelov, A.V. Fomin, A.A. Volkov, V.K. Kotov, *The theory of signal transmission in rail transport* (M, Transport, 2001).
5. V.L. Lazarev, Entropy approach to the organization of monitoring and management *Proceedings of the Russian Academy of Sciences, Theory and Control Systems* **6**, 61-68 (2005).
6. M.V. Prangvishvili, *Entropic and other systemic laws: Management issues of complex systems* (Moscow, Nauka, Institute of Management Problems Trapeznikova, 2003).
7. V.G. Polosin, S.V. Tertychnaya, News of higher educational institutions. Volga region. Technical sciences **1(9)**, 127-133 (2009).
8. O.V. Mansurov, *Equation's math physics. The theory of functions of a complex variable. Numerical methods. Probability theory* (Moscow, Higher school, 1991).
9. G.I. Ivchenko, Yu.I. Medvedev, *Introduction to mathematical statistics* (Moscow, Publishing house LCI, 2010).
10. C.E. Shannon, *The mathematical theory of communication. Work on information theory and cybernetics* (Moscow, Publishing House of Foreign Literature, 1963).
11. E.S. Wenzel, *Probability theory* (Moscow, Higher school, 1998).
12. V.E. Kraus, *Measuring information systems* (Moscow, Mir Publishing House, 1975).