DOI 10.1007/s11018-022-01990-0
Measurement Techniques, Vol. 64, No. 9, December, 2021

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

FORMATION OF SETS OF INDEPENDENT COMPONENTS OF A MULTIDIMENSIONAL RANDOM VARIABLE BASED ON A NONPARAMETRIC PATTERN RECOGNITION ALGORITHM

A. V. Lapko,[1,2] V. A. Lapko,[1,2] and A. V. Bakhtina[2]

UDC 519.7+004.93

We consider the possibility of circumventing the decomposition problem for the range of values of random variables when testing various hypotheses. A brief review of the literature on this issue is given. A method is proposed for forming sets of independent components of a multidimensional random variable, based on testing hypotheses about the independence of paired combinations of components of a multidimensional random variable. The method uses a two-dimensional nonparametric algorithm for the recognition of kernel-type patterns, corresponding to the criterion of maximum likelihood. In contrast to the traditional technique using Pearson's criterion, the proposed technique avoids the problem of decomposing the range of values of random variables into multidimensional intervals. We present results of computational experiments performed using the method of forming sets of independent random variables. From the obtained data, an information graph is constructed, whose vertices correspond to the components of a multidimensional random variable, and the edges determine their independence, while the vertices of the complete subgraphs correspond to groups of independent components of the random variable. The results obtained form the basis for the synthesis of a multilevel nonparametric system for processing large volumes of data, each level of which corresponds to a specific set of independent random variables.

Keywords: *hypothesis testing, a set of independent random variables, multidimensional random variable, pattern recognition algorithms, kernel estimate of probability density, choice of blur coefficients of kernel functions, counter-excess coefficient, asymmetry coefficient, information graph.*

Introduction. The formation of sets of independent random variables is necessary when reducing the dimension of information processing problems and the synthesis of effective decision-making algorithms. A priori information about the independence of random variables makes it possible to improve the approximation properties of the nonparametric estimate of the probability density in comparison with the kernel statistics for dependent variables [1–3]. The results obtained in these works are confirmed by studying the asymptotic properties of a nonparametric estimate for the equation of a separating surface of kernel type in the problem of pattern recognition [4].

The traditional method for testing the hypothesis of the independence of random variables is based on the Pearson χ^2 test. However, its application includes the difficultly formalized stage of dividing the range of values of random variables into multidimensional intervals and requires large volumes of initial statistical data, which is associated with the transition from continuous to discrete random variables [5]. Methods for discretizing the interval of values of a one-dimensional random variable are

[1] Institute of Computational Modeling, Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia; e-mail: lapko@icm.krasn.ru.
[2] Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia; e-mail: anna-denisyuk@yandex.ru.

Translated from Izmeritel'naya Tekhnika, No. 9, pp. 3–9, September, 2021. Original article submitted February 17, 2021. Accepted May 18, 2021.

considered in [6–8]. In [7, 9], formulas for the discretization of the range of values of a multidimensional random variable with a normal distribution law are given, obtained on the basis of an analysis of the asymptotic properties of the histogram. The work [10] substantiates the method of optimal discretization of the range of values of a multidimensional random variable. This technique is consistent with the formula for choosing the number of sampling intervals for a one-dimensional random variable with a uniform distribution law (Heyhold and Gaede's formula) [11], and its implementation is coupled with the estimation of the integral of the square of a multidimensional probability density. When evaluating this functional on the probability density, the first positive results were obtained [12, 13]. Therefore, it was required to develop a new method for testing the hypothesis under consideration, which avoids the decomposition problem for the domain of values of random variables. A similar problem is solved when testing the hypothesis about the identity of the distribution laws of random variables based on a nonparametric pattern recognition algorithm [14–16]. These works show the possibility of replacing the hypothesis testing problem about the distributions of random variables with the hypothesis testing problem about the equality of the pattern recognition error to a certain threshold value.

The purpose of this article is to develop a methodology for the formation of sets of independent components of a multidimensional random variable using a nonparametric algorithm for pattern recognition of kernel type that meets the criterion of maximum likelihood.

Testing the hypothesis of the independence of the components of a two-dimensional random variable. Let there be a sample $V = (x_1^i, x_2^i, i = 1, ..., n)$ of volume n formed from independent observations of a two-dimensional random variable $x = (x_1, x_2)$. Observations x are extracted from general populations characterized by unknown probability densities $p(x_1)p(x_2)$ or $p(x_1, x_2)$. It is necessary to check the hypothesis of the independence of the random variables x_1, x_2:

$$H_0\colon p(x_1, x_2) \equiv p(x_1)p(x_2). \tag{1}$$

To test the hypothesis H_0 (1), we will solve a two-alternative pattern recognition problem. The classes Ω_1, Ω_2 correspond to the domains of definition of the probability densities $p(x_1)p(x_2)$, $p(x_1, x_2)$. Under these conditions, the Bayesian decision rule corresponding to the criterion of maximum likelihood has the form

$$m(x)\colon \begin{cases} x \in \Omega_1 \;\text{ if }\; p(x_1, x_2) < p(x_1)p(x_2); \\ x \in \Omega_2 \;\text{ if }\; p(x_1, x_2) > p(x_1)p(x_2). \end{cases} \tag{2}$$

In contrast to the traditional formulation of the pattern recognition problem, in the synthesis of the decision rule (2) under conditions of initial uncertainty, there is clearly no training sample. The estimation of the probability densities $p(x_1)p(x_2)$, $p(x_1, x_2)$ is carried out using the sample V. For this, nonparametric estimates of the probability densities of the Rosenblatt–Parzen type are used [1–3, 17, 18]:

$$\bar{p}(x_1)\bar{p}(x_2) = \frac{1}{n^2 c_1 c_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^j}{c_2}\right); \tag{3}$$

$$\bar{p}(x_1, x_2) = \frac{1}{n c_1 c_2} \sum_{i=1}^{n} \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^i}{c_2}\right), \tag{4}$$

where c_1, c_2 are the blur coefficients of the kernel functions $\Phi(u_v)$, $v = 1, 2$.

In statistics (3) and (4), the kernel functions $\Phi(u_v)$ satisfy the conditions

$$0 \le \Phi(u_v) < \infty; \quad \Phi(u_v) = \Phi(-u_v); \quad \int_{-\infty}^{+\infty} \Phi(u_v)\, du_v = 1; \quad \int_{-\infty}^{+\infty} u^m \Phi(u_v)\, du_v < \infty; \quad 0 \le m < \infty; \quad v = 1, 2.$$

The values of the kernel function blur coefficients c_1, c_2 decrease with an increase in the sample volume n of statistical data V.

The nonparametric statistics (3), (4) are asymptoticly unbiased and consistent [1, 2, 18]. It was found that the minimum value of the standard deviation (RMSD) $\bar{p}(x_1)\bar{p}(x_2)$ from $p(x_1)p(x_2)$ with an increase in the volume n of initial statistical data tends to zero in proportion to the value $n^{-4/5}$. The order of such convergence of the nonparametric estimate $\bar{p}(x_1, x_2)$ to the probability density $p(x_1, x_2)$ is less and amounts to $n^{-4/6}$. The a priori information on the independence of random variables enables an increase in the order of convergence of a nonparametric kernel-type probability density estimate.

Taking into account expressions (2)–(4), we write the nonparametric decision rule for the classification of random variables $x = (x_1, x_2)$ as

$$\bar{m}(x): \begin{cases} x \in \Omega_1 \ \text{ if } \ \bar{p}(x_1, x_2) < \bar{p}(x_1)\bar{p}(x_2); \\ x \in \Omega_2 \ \text{ if } \ \bar{p}(x_1, x_2) > \bar{p}(x_1)\bar{p}(x_2). \end{cases} \tag{5}$$

In the modification of the nonparametric algorithm for pattern recognition (5), the optimal blur coefficients c_1, c_2 of the kernel estimates of the probability densities $\bar{p}(x_1)\bar{p}(x_2)$ and $\bar{p}(x_1, x_2)$ are selected based on the results of the analysis of their approximation properties.

The optimal value of the blur coefficient c_v of the kernel functions of the nonparametric estimate of the one-dimensional probability density under the condition of the minimum standard deviation $\bar{p}(x_v)$ of $p(x_v)$ is determined by the equation [18]:

$$c_v^* = [\|\Phi(u)\|^2/(n\|p^{(2)}(x_v)\|^2)]^{1/5}. \tag{6}$$

Here $p^{(2)}(x_v)$ is the second derivative of the probability density $p(x_v)$ with respect to x_v;

$$\|\Phi(u)\|^2 = \int_{-\infty}^{\infty} \Phi^2(u)\,du; \quad \left\|p^{(2)}(x_v)\right\|^2 = \int_{-\infty}^{\infty} \left(p^{(2)}(x_v)\right)^2 dx_v.$$

After simple transformations, we represent Eq. (6) in the form

$$c_v^* = \beta_v \sigma_v n^{-1/5},$$

where σ_v is the standard deviation of the random value x_v;

$$\beta_v = [\|\Phi(u)\|^2/(\sigma_v^5\|p^{(2)}(x_v)\|^2)]^{1/5}.$$

The component of the functional β_v, defined by the expression

$$\lambda_v = \sigma_v^5\|p^{(2)}(x_v)\|^2,$$

is a constant for a number of unimodal probability densities. The value λ_v is determined by the form of the probability density and does not depend on its parameters [19–27].

According to the data of computational experiments for a family of one-dimensional lognormal distribution laws, the authors of this article have determined the parameter estimates

$$\bar{\beta}_v = 1.49\exp(-0.589\bar{\alpha}_v^{0.857}) + 0.021,$$

where $\bar{\alpha}_v = (\bar{\delta}_v^2 + \bar{\eta}_v^2)^{1/2}$; $\bar{\delta}_v$ and $\bar{\eta}_v$ are the estimates of the coefficients of counter-excess and asymmetry of the random variable x_v, $v = 1, 2$, respectively.

The optimal kernel function blur coefficient in the statistic $\bar{p}(x_v)$ is estimated by the formula

$$\bar{c}_v^* = \bar{\beta}_v \bar{\sigma}_v n^{-1/5}, v = 1, 2. \tag{7}$$

According to the proposed methodology and taking into account the results of research [27], for a two-dimensional random variable $x = (x_1, x_2)$, the estimates of the optimal blur coefficients of the kernel statistic $\bar{p}(x_1, x_2)$ are determined by the expression

$$\bar{\bar{c}}_v^* = \bar{\beta}\bar{\sigma}_v n^{-1/6}, v = 1, 2. \tag{8}$$

In Eq. (8), the parameter $\bar{\beta}$ takes on the value

$$\bar{\beta} = 1.498\exp(-0.524\bar{\alpha}^{0.809}) + 0.0356,$$

where $\bar{\alpha} = (\bar{\delta}_1^2 + \bar{\delta}_2^2 + \bar{\eta}_1^2 + \bar{\eta}_2^2)^{1/2}$.

The obtained estimates for the functionals β_v, $v = 1, 2$, and β refine the research results [27–29].

Let us estimate the error probabilities of recognizing ρ_1, ρ_2 using the nonparametric decision rule (5) for pattern recognition from the statistical data V. In this case, we will take into account the estimates of the blur coefficients (7), (8) of the kernel statistics $\bar{p}(x_1)\bar{p}(x_2)$, $\bar{p}(x_1, x_2)$.

The estimates $\bar{\rho}_t$ are calculated in the “sliding exam” mode on the sample V under the assumption that its elements belong to the class Ω_t:

$$\bar{\rho}_t = \frac{1}{n}\sum_{j=1}^{n} 1\left(\delta(j), \bar{\delta}(j)\right),$$

where $1(\delta(j), \bar{\delta}(j))$ is the indicator function; $\delta(j) = t$ is an indicator of type $x^j = (x_1^j, x_2^j) \in \Omega_t$; $\bar{\delta}(j)$ is the “decision” of algorithm (5) for membership of the situation x^j in one of the classes Ω_t, $t = 1, 2$.

When calculating $\bar{\rho}_t$ using the “sliding exam” methodology, the situation $x^j = (x_1^j, x_2^j)$ from the sample V, which is controlled by algorithm (5), is excluded from the process of generating statistics (3) and (4).

When forming the values $\bar{\rho}_t$, the indicator function is defined as

$$1\left(\delta(j), \bar{\delta}(j)\right) = \begin{cases} 0 \ \text{ if } \ \delta(j) = \bar{\delta}(j); \\ 1 \ \text{ if } \ \delta(j) \neq \bar{\delta}(j). \end{cases}$$

Let us compare the values $\bar{\rho}_1, \bar{\rho}_2$ under the assumption that the elements of the sample V belong to the classes Ω_1, Ω_2, respectively. The hypothesis H_0 is satisfied if $\bar{\rho}_1 < \bar{\rho}_2$. This assertion is true, since with the independence of the random variables in the domains of definition Ω_1, Ω_2 of the estimates of the probability densities $\bar{p}(x_1)\bar{p}(x_2)$ and $\bar{p}(x_1, x_2)$, the relation $\bar{p}(x_1)\bar{p}(x_2) >$ $> \bar{p}(x_1, x_2)$ is satisfied with the estimated error probability $\bar{\rho}_1$. Otherwise, if $\bar{\rho}_2 < \bar{\rho}_1$, the random variables x_1, x_2 are dependent.

The probabilities of Bayesian errors ρ_1, ρ_2 of class recognition are linear functionals of the probability densities $p(x_1)p(x_2)$ and $p(x_1, x_2)$, respectively. Since the nonparametric estimates of the indicated probability densities have the properties of asymptotic convergence [4, 18], the asymptotic convergence of the statistical estimates $\bar{\rho}_1, \bar{\rho}_2$ to ρ_1, ρ_2 is assumed.

With bounded volumes n of the initial sample V, the problem of confidence-based estimation of the probabilities of pattern recognition errors arises. To solve it, one can use the traditional method of confidence assessment of probabilities [5] or the Kolmogorov–Smirnov test [30], in which the deviation $\bar{D}_{12} = |\bar{\rho}_1 - \bar{\rho}_2|$ is compared with the threshold value $D_\beta = [-\ln(\beta_0/2)/n]^{1/2}$. Here β_0 is the probability (risk) of rejecting the hypothesis $\bar{H}_0$: $\rho_1 < \rho_2$. If the relation $\bar{D}_{12} < D_\beta$ holds, then the hypothesis $\bar{H}_0$ is valid and the risk of rejecting it does not exceed the value β_0. When $\bar{D}_{12} > D_\beta$, the hypothesis $\bar{H}_0$ is rejected.

Methods for the formation of sets of components of a multidimensional random variable. Suppose there is a sample of observations $V = (x_v^i, v = 1, ..., k, i = 1, ..., n)$ of volume n, composed of statistically independent observations of the components of a multidimensional random variable $x = (x_v, v = 1, ..., k)$. The form of the probability density $p(x)$ is a priori unknown. From the statistical data of the sample V, using the above proposed hypothesis testing criterion

$$H_{vj}: p(x_v, x_j) \equiv p(x_v)p(x_j) \tag{9}$$

for the components x_v, $v = 1, ..., k$, x_j, $j = 1, ..., k$, $v > j$, it is required to generate sets of independent random variables $x(t) =$ $= (x_v, v \in I_t)$, $t = 1, ..., m$. Here I_t is the set of indices of components that make up the set $x(t)$, and the number m of sets of components of a random variable x is unknown.

The proposed technique consists of three stages.

Stage 1. In accordance with the above recommendations, the hypothesis H_{vj} of type (9) is tested for each pair of components (x_v, x_j) of a multidimensional random variable $x = (x_v, v = 1, ..., k)$. The number of such pairs is $0.5k(k - 1)$.

Stage 2. Based on the results of stage 1, an information graph $G(X, A)$ is constructed, where X is the set of vertices corresponding to the components of the random variable x; A is the set of edges. If the hypothesis H_{vj} holds, that is, the components x_v, x_j are independent, then there is an edge between two vertices x_v, x_j.

Stage 3. Analyze the information graph $G(X, A)$ and determine the complete subgraphs $G(X_t, A_t)$, $t = 1, ..., m$. If the components of the random variable x are independent, then each pair of vertices of the subgraph $G(X_t, A_t)$ has an edge. Detect the complete subgraphs with algorithms to decompose the original graph using its adjacency matrix [31]. The components x_v, $v \in I_t$, corresponding to the vertices of the complete subgraph $G(X_t, A_t)$ form a set of independent random variables. In this case, one can find a number of options for decomposing the information graph.

Analysis of the results of computational experiments. Let us check the efficiency of the proposed method when analyzing the data of a computational experiment. Let us investigate the efficiency of the procedure for forming sets of independent components on the volume n of the initial statistical data and the degree of dependence of random variables.

TABLE 1. Results of Testing Hypotheses about the Independence of Paired Combinations of a Four-Dimensional Random Variable

n	Parameter	Equation	x_1, x_2	x_1, x_3	x_1, x_4	x_2, x_3	x_2, x_4	x_3, x_4
100	$\bar{\rho}_1$	(10)	0.51	0.52	0.75	0.59	0.57	0.59
		(11)	0.52	0.56	0.77	0.67	0.76	0.65
	$\bar{\rho}_2$	(10)	0.49	0.48	0.25	0.41	0.43	0.41
		(11)	0.48	0.44	0.23	0.33	0.24	0.35
	$\bar{D}$	(10)	0.02	0.04	0.50	0.18	0.14	0.18
		(11)	0.04	0.12	0.54	0.34	0.52	0.30
500	$\bar{\rho}_1$	(10)	0.460	0.432	0.808	0.48	0.498	0.442
		(11)	0.448	0.476	0.706	0.49	0.822	0.430
	$\bar{\rho}_2$	(10)	0.540	0.568	0.192	0.52	0.502	0.558
		(11)	0.552	0.524	0.294	0.51	0.178	0.570
	$\bar{D}$	(10)	0.080	0.136	0.616	0.04	0.004	0.116
		(11)	0.104	0.048	0.412	0.020	0.644	0.140
1000	$\bar{\rho}_1$	(10)	0.422	0.462	0.789	0.479	0.473	0.451
		(11)	0.495	0.475	0.725	0.458	0.813	0.480
	$\bar{\rho}_2$	(10)	0.578	0.538	0.211	0.521	0.527	0.549
		(11)	0.505	0.525	0.275	0.542	0.187	0.520
	$\bar{D}$	(10)	0.156	0.076	0.578	0.042	0.054	0.098
		(11)	0.010	0.050	0.450	0.084	0.626	0.040

When designing computational experiments, the statistical data $V = (x_1^i, ..., x_4^i, i = 1, ..., n)$ of the components x_1, x_2, x_3 of a multidimensional random variable $x = (x_1, x_2, x_3, x_4)$ are assumed to be independent. Their values are formed using sensors with uniform $p(x_1) = R(3;\, 1)$ and normal distribution laws $p(x_2) = N(3;\, 1), p(x_3) = N(3;\, 1)$, where $R(3;\, 1)$ and $N(3;\, 1)$ are the distribution laws of random variables with mathematical expectation and standard deviation equal to three and one, respectively. The values of the component x_4 are found from one or another dependence determined by various conditions of the study:

$$x_4 = \varphi(x_1) = x_1^2 - 6x_1 + 10 + \varepsilon; \tag{10}$$

$$x_4 = \varphi_4(x_1, x_2) = x_1^2 - x_2^2 + 6(x_2 - x_1) + 20 + \varepsilon. \tag{11}$$

Here ε are the values of a random variable with a normal distribution law $N(0;\, 1)$.

Sensors of random variables x_2, x_3 with normal distribution laws are formed on the basis of expressions

$$x_\nu = 3 + \left(\sum_{l=1}^{r} \varepsilon_0^l - 0.5r \right) 6 \big/ \sqrt{3r}\,, \quad \nu = 2, 3,$$

where ε_0^l are values of random variables with a uniform probability density on the interval $[0;\, 1]$; $r = 12$.

The technique of computational experiments is implemented in the Delphi-7 programming environment. To generate a random variable $\varepsilon_0 \in (0;\, 1)$ with a uniform distribution law, we use a standard function *random* and the procedure Randomize, which takes into account the time of day as the basis for generating pseudo-random numbers with a uniform distribution law.

The volume n of initial statistical data in computational experiments was 100, 500, 1000. For a specific volume n of initial data, the parameters $\bar{\rho}_1$, $\bar{\rho}_2$, $\bar{D}$ correspond to dependencies (10), (11). The parameter values $\bar{\rho}_1$, $\bar{\rho}_2$ are calculated 10 times and then averaged (cf. Table 1).

With relatively small volumes $n = 100$ of initial statistical data and using dependence (10), the values $\bar{\rho}_1$ and $\bar{\rho}_2$ for pairs of random variables (x_1, x_2), (x_1, x_3), (x_2, x_3), (x_2, x_4) and (x_3, x_4), are not reliably different. Under these conditions, the

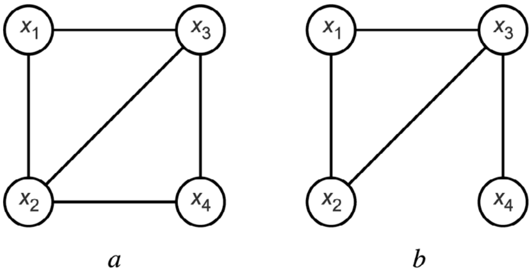

Fig. 1. Information graphs $G(X, A)$ for $n = 1000$ using transformation (10), (11) (a, b, respectively).

values $\bar{D}_{12}, \bar{D}_{13}, \bar{D}_{23}, \bar{D}_{24}, \bar{D}_{34}$, are less than the threshold $D_\beta = 0.192$ with the risk $\beta_0 = 0.05$ of rejecting the hypothesis $\bar{H}_0$. According to the conditions of the computational experiment, these paired sets of components of the random variable are a priori independent. With small volumes of initial data, there is an uncertainty in decision making, which follows from the analysis of the results of testing the hypotheses under study. This conclusion is explained by the difference in the convergence conditions, for example, in the nonparametric estimates of the probability densities $\bar{p}(x_1)\bar{p}(x_2)$ and $\bar{p}(x_1, x_2)$, which is confirmed by the results of studies in [1–3].

If the paired combinations of random components are dependent, then the proposed method, under the conditions of the considered computational experiment, unambiguously rejects the hypothesis $\bar{H}_0$.This conclusion is valid for the components (x_1, x_4) when using dependence (10) in the computational experiment. In this case, the inequality $\bar{\rho}_1 > \bar{\rho}_2$ holds, i.e., the condition $\bar{p}(x_1, x_2) > \bar{p}(x_1)\bar{p}(x_2)$ is satisfied, and the decision that is made according to rule (5) is valid. The specified condition is reliable, since $\bar{D}_{14} < D_\beta$ when $\bar{D}_{14} = 0.5$ and $D_\beta = 0.192$. With the complication of dependencies between random variables under the conditions of using the transformation (11) with $n = 100$, the above-mentioned regularities are preserved.

With an increase in the volume n of the initial statistical data V, the efficiency of the proposed method for testing hypotheses about the independence of random variables increases. With the volume of initial data $n = 500$, the hypothesis about the independence of paired combinations of random variables (x_1, x_2), (x_1, x_3), (x_2, x_3), (x_3, x_4) is fulfilled, since the corresponding errors of pattern recognition by the decision rule (5) are connected by the relation $\bar{\rho}_1 < \bar{\rho}_2$. This relation holds for the paired combinations (x_1, x_3), (x_3, x_4) when using dependence (10) and the risk $\beta_0 = 0.05$ in the computational experiment to reject the hypothesis $\bar{H}_0$. The dependence between the components (x_1, x_4) is more significant when $\bar{D}_{14} = 0.616$ and $D_\beta = 0.086$. As the relationship between random variables (11) becomes more complex, a reliable dependence is observed between (x_1, x_4), (x_2, x_4), as well as a reliable independence of the components (x_1, x_2), (x_3, x_4).

With the volume of initial data n =1000, the above-mentioned patterns for $n = 500$ are basically preserved. Application of the developed methodology enables detection of dependency characteristics contradicting the hypothesis $\bar{H}_0$ with a high level of reliability when using transformations (10), (11) in a computational experiment. If the components of the random variable x are a priori independent, then the inequality $\bar{\rho}_1 < \bar{\rho}_2$ holds; however, under the conditions of the considered computational experiment, it can be reliable or unreliable.

Based on the data in the Table 1, we will form an information graph $G(X, A)$, where X is the set of vertices that correspond to the components x_v, $v = 1, ..., 4$, of the random variable x; A is the set of edges between the vertices of the graph. There is an edge between two vertices x_v, x_j if the corresponding components are independent (cf. Fig. 1).

When using transformation (10), there are two versions of complete subgraphs, which correspond to sets of independent components (x_1, x_2, x_3) and (x_2, x_3, x_4) (cf. Fig. 1*a*). The choice of a specific option is determined by the method of subsequent processing of the initial data. If transformation (11) is applied, then one set of independent random variables (x_1, x_2, x_3) is found (cf. Fig. 1*b*). The results obtained correspond to the conditions for designing a computational experiment based on transformations (10), (11).

Conclusion. The proposed method for the formation of sets of independent components of a multidimensional random variable can be replaced by testing hypotheses about the independence of their paired combinations. To solve this problem, a two-dimensional nonparametric algorithm of kernel-type pattern recognition is used, which corresponds to the criterion of maximum likelihood. This approach allows us to bypass the problem of decomposition of the values of random variables into multidimensional intervals. When optimizing a nonparametric pattern recognition algorithm, it is advisable to use fast algorithms for selecting the blur coefficients, which is especially important when processing large volumes of data. Based on the results of testing hypotheses about the independence of two-dimensional random variables it is possible to build an information graph, whose vertices correspond to random variables, and edges define their independence. Decomposition of the information graph into complete subgraphs enables detection of different variants of sets of independent components of a multidimensional random variable. The choice of a particular set is determined by the adopted procedure for the subsequent processing of the initial data.

According to the results of computational experiment, it was established that the proposed technique is especially sensitive to the detection of dependent random variables, which is characteristic of small and large volumes of initial statistical data. In conditions of small volumes of values of a four-dimensional random variable (for n <500), it is impossible to make an unambiguous decision about the independence of random variables. At $n \geq 500$ under the conditions of a computational experiment, it is possible to reliably detect the dependent components of a four-dimensional random variable. Independent random variables in these conditions are determined with varying degrees of reliability with limited amounts of initial statistical data. A promising direction for further research is the comparison of the proposed method with the traditional one based on the use of Pearson's criterion with different formulas for discretization of the range of values of random variables.

The results obtained form the basis for the synthesis of a multi-level nonparametric system, where each level corresponds to a specific set of independent random variables. Such systems are efficient when processing large data volumes.

The study was carried out with the financial support of the Russian Foundation for Basic Research, the Government of the Krasnoyarsk Territory, and the Krasnoyarsk Regional Science Foundation within the framework of the scientific project No. 20-41-240001.

REFERENCES

1. A. V. Lapko and V. A. Lapko, "Properties of a nonparametric estimate of the multidimensional probability density of independent random variables," *Informat. Sist. Upravl.*, **31**, No. 1, 166–174 (2012).
2. A. V. Lapko and V. A. Lapko, "Nonparametric estimate of the probability density of independent random variables," *Informat. Sist. Upravl.*, **29**, No. 3, 118–124 (2011).
3. A. V. Lapko and V. A. Lapko, "Influence of a priori information about the independence of multidimensional random variables on the properties of their nonparametric estimate of the probability density," *Sist. Upravl. Inform. Tekhnol.*, **48**, No. 2.1, 164–167 (2012).
4. A. V. Lapko and V. A. Lapko, "Properties of a nonparametric decision function in the presence of a priori information about the non-dependence of the attributes of classified objects," *Avtometriya*, **48**, No. 4, 112–119 (2012).
5. V. S. Pugachev, *Probability Theory and Mathematical Statistics: Textbook*, Fizmatlit, Moscow (2002).
6. H. A. Sturges, *J. Am. Stat. Ass.*, **21**, pp. 65–66 (1926), https://doi.org/10.1080/01621459.1926.10502161.
7. D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York (1992).
8. A. Hacine-Gharbi, P. Ravier, R. Harba, and T. Mohamadi, *Patt. Recogn. Lett.*, **33**, No. 10, 1302–1308 (2012), https://doi.org/10.1016/j.patrec.2012.02.022.
9. L. Devroye and G. Lugosi, *Test*, **13**, No. 1, 129–145 (2004), https://doi.org/10.1007/ BF02603004.
10. A. V. Lapko and V. A. Lapko, "Method of discretization of the region of values of a multidimensional random variable," *Izmerit. Tekhn.*, No. 1, 16–20 (2019), https://doi.org/10.32446/0368-1025it.2019-1-16-20.
11. I. Heinhold and K. Gaede, *Ingeniur statistic*, Springler Verlag, München-Wien (1964).
12. A. V. Lapko and V. A. Lapko, "Nonparametric estimation of the quadratic functional of multimodal probability density," *Metrologiya*, No. 3, 17–29 (2019), https://doi.org/10.32446/0132-4713.2019-3-17-29.
13. A. V. Lapko and V. A. Lapko, "Estimation of the parameters of the formula for optimal discretization of the range of values of a two-dimensional random variable," *Izmer. Tekhn.*, No. 5, 9–13, (2018), https://doi.org/10.32446/0368-1025it.2018-5-9-13.

14. A. V. Lapko and V. A. Lapko, “Nonparametric algorithms for pattern recognition in the problem of testing the statistical hypothesis of the identity of two distribution laws of random variables,” *Avtometriya*, **46**, No. 6, 47–53 (2010).

15. A. V. Lapko and V. A. Lapko, “Comparison of empirical and theoretical distribution functions of a random variable on the basis of a nonparametric classifier,” *Avtometriya*, **48**, No. 1, 45–49 (2012).

16. A. V. Lapko and V. A. Lapko, “Technique for testing hypotheses about the distributions of multidimensional spectral data using a nonparametric pattern recognition algorithm,” *Komp. Opt.*, **43**, No. 2, 238–244 (2019), https://doi.org/10.18287/2412-6179-2019-43-2-238-244.

17. E. Parzen, *Ann. Math. Stat.*, **33**, No. 3, 1065–1076 (1962).

18. V. A. Epanechnikov, “Nonparametric estimate of multidimensional probability density,” *Teor. Veroyatn. Primen.*, **14**, No. 1, 156–161 (1969).

19. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London (1986).

20. S. Sheather and M. Jones, *J. Roy. Stat. Soc. B*, **53**, No. 3, 683–690 (1991), https://doi.org/10.1111/j.2517-6161.1991.tb01857.x.

21. S. J. Sheather, *Stat. Sci.*, **19**, No. 4, 588–597 (2004), https://doi.org/10.1214/088342304000000297.

22. G. R. Terrell and D. W. Scott, *J. Am. Stat. Ass.*, **80**, 209–214 (1985).

23. M. C. Jones, J. S. Marron, and S. J. Sheather, *J. Am. Stat. Ass.*, **91**, 401–407 (1996).

24. D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, New Jersey (2015).

25. A. V. Lapko and V. A. Lapko, “Modified algorithm for fast selection of the blur coefficients of nuclear estimates of multidimensional probability densities,” *Izmer. Tekhn.*, No.11, 9–13, (2020), https://doi.org/10.32446/0368-1025it.2020-11-9-13.

26. A. V. Lapko and V. A. Lapko, “Estimation of the integral of the square of derivatives of symmetric probability densities of one-dimensional random variables,” *Metrologiya*, No. 1, 15–27 (2020), https://doi.org/10.32446/0132-4713.2020-1-15-27.

27. A. V. Lapko and V. A. Lapko, “Estimation of a nonlinear functional of probability density for the optimization of nonparametric decision functions,” *Izmer. Tekhn.*, No. 1, 14–20 (2021), https://doi.org/10.32446/0368-1025it.2021-1-14-20.

28. A. V. Lapko and V. A. Lapko, “A fast algorithm for choosing kernel function blur coefficients in a nonparametric estimate of the probability density,” *Izmer. Tekhn.*, No. 6, 16–20 (2018). 2018, https://doi.org/10.32446/0368-1025it-2018-6-16-20.

29. A. V. Lapko and V. A. Lapko, “A fast algorithm for the selection of blur coefficients in multidimensional kernel estimates of the probability density,” *Izmer. Tekhn.*, No. 10, 19–23 (2018), https://doi.org/10.32446/0368-1025it.2018-10-19-23.

30. A. S. Sharakshane, I. G. Zheleznov, and V. A. Ivnitskiy, *Complex Systems*, Vysshaya Shkola, Moscow (1977).

31. N. Christofides, *Graph Theory: an Algorithmic Approach* [Russian translation], Mir, Moscow (1978).