

Testing the Hypothesis of the Independence of Two-Dimensional Random Variables Using a Nonparametric Algorithm for Pattern Recognition

A. V. Lapko^{1,2*} and V. A. Lapko^{1,2}

¹*Institute of Computational Modelling, Siberian Branch, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*

²*Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, 660037 Russia*

Received December 17, 2020; revised March 4, 2021; accepted March 4, 2021

Abstract—A new method for testing the hypothesis of the independence of two-dimensional random variables is proposed. The method under consideration is based on the use of a nonparametric algorithm for pattern recognition that meets the maximum likelihood criterion. In contrast to the traditional problem statement, there is no training sample a priori. The initial information is represented by statistical data that make up the values of two-dimensional random variables. The laws of distribution of random variables in classes are estimated from the initial statistical data for the conditions of their dependence and independence. When choosing the optimal blur coefficients for nonparametric estimates of probability densities, the maximum of the likelihood functions is used as a criterion. Under these conditions, estimates of the probability of pattern recognition errors in classes are calculated. Based on the minimum value of the estimate of the probability of an error in pattern recognition, a decision is made on the independence or dependence of random variables. The effectiveness of the developed method is confirmed by the results of computational experiments when testing the hypothesis of the independence or linear dependence of two-dimensional random variables.

DOI: 10.3103/S8756699021020114

Keywords: testing the hypothesis of the independence of random variables, two-dimensional random variables, pattern recognition, kernel probability density estimation, maximum-likelihood criterion, confidence estimation of probabilities

INTRODUCTION

Information about the dependence or independence of random variables is a necessary condition for the synthesis of effective algorithms for information processing and decision-making. In [1–3], the properties of a nonparametric Rosenblatt–Parzen-type probability density estimation of independent random variables are investigated. It is found that the presence of a priori information about the independence of random variables allows us to improve the approximation properties of the nonparametric estimation of their probability density in comparison with the kernel statistics for dependent random variables. This advantage increases with the increase in the dimension of random variables. The obtained results are confirmed by studying the asymptotic properties of the nonparametric estimation of the separating surface equation in the two-alternative problem of pattern recognition [4].

The traditional method of testing the hypothesis of the independence of random variables is based on the use of the universal χ^2 -Pearson criterion. However, its formation contains a difficult-to-formalize stage of dividing the range of values of random variables into multidimensional intervals [5]. Therefore, the problem arises of developing a new method of testing the hypothesis providing a bypass for the problem of decomposing the domain of values of random variables. A similar problem is solved when

*E-mail: lapko@icm.krasn.ru

testing the hypothesis of the identity of the distribution laws of random variables based on the use of a nonparametric pattern recognition algorithm [6–8]. It is shown that it can be replaced by the task of testing the hypothesis that the image recognition error is equal to a certain threshold value. The training sample in the synthesis of a nonparametric pattern recognition algorithm is formed based on statistical data that characterize the distribution laws of the compared random variables.

The purpose of this paper is to develop the proposed approach to the problem of testing the hypothesis of the independence of random variables using a nonparametric pattern recognition algorithm.

METHOD FOR TESTING THE HYPOTHESIS OF THE INDEPENDENCE OF RANDOM VARIABLES

Let there be a sample $V = (x^i, i = \overline{1, n})$ of volume n made up of independent observations of a two-dimensional random variable $x = (x_1, x_2)$. The sample V is extracted from general populations characterized by probability densities $p(x_1)p(x_2)$ or $p(x_1, x_2)$. It is necessary to check the hypothesis

$$H_0: p(x_1, x_2) \equiv p(x_1)p(x_2) \quad (1)$$

about the independence of random variables x_1, x_2 using the statistical data V .

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

$$m(x): \begin{cases} x \in \Omega_1, & \text{of } p(x_1, x_2) < p(x_1)p(x_2); \\ x \in \Omega_2, & \text{of } p(x_1, x_2) > p(x_1)p(x_2). \end{cases} \quad (2)$$

In contrast to the traditional formulation of the pattern recognition problem, the synthesis of the decision rule (2) a priori lacks a training sample containing information about the membership of elements of the sample V to a particular class. This information should be detected during the implementation of the H_0 hypothesis testing technique, which is based on performing the following actions.

From the sample V , we reconstruct the probability densities $p(x_1, x_2), p(x_1)p(x_2)$ using their nonparametric Rosenblatt–Parzen-type estimates [9, 10]:

$$\bar{p}(x_1, x_2) = \frac{1}{nc_1c_2} \sum_{i=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^i}{c_2}\right), \quad (3)$$

$$\bar{p}(x_1)\bar{p}(x_2) = \frac{1}{n^2c_1c_2} \sum_{i=1}^n \sum_{j=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^j}{c_2}\right). \quad (4)$$

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

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

The values of the blur coefficients c_v of the kernel functions decrease with the growth of the volume n of the statistical data sample V . Taking into account expressions (2)–(4), we can write the nonparametric decision rule for the classification of random variables $x = (x_1, x_2)$ as

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

Under conditions of such uncertainty, we will choose the optimal blur coefficients of the kernel functions of the decision rule (5) based on the approximation properties for the nonparametric estimates $\bar{p}(x_1, x_2), \bar{p}(x_1)\bar{p}(x_2)$ of probability densities $p(x_1, x_2), p(x_1)p(x_2)$. To select the optimal blur coefficients of the nonparametric estimate of the probability density $p(x_1, x_2)$, as a criterion, for example, the maximum of the likelihood function is used [11, 12]

$$L(c_1, c_2) = \prod_{j=1}^n \bar{p}(x_1^j, x_2^j), \quad \bar{p}(x_1^j, x_2^j) = \frac{1}{(n-1)c_1c_2} \sum_{i=1, i \neq j}^n \Phi\left(\frac{x_1^j - x_1^i}{c_1}\right) \Phi\left(\frac{x_2^j - x_2^i}{c_2}\right). \quad (6)$$

By analogy with expression (6), it is easy to determine the criterion for choosing the optimal blur coefficients for the statistics $\bar{p}(x_1)\bar{p}(x_2)$ (4).

Note that the choice of the optimal blur coefficients of nonparametric estimates of the probability densities $\bar{p}(x_1, x_2), \bar{p}(x_1)\bar{p}(x_2)$ can be carried out from the condition of the minimum statistical estimates of the standard deviations $\bar{p}(x_1, x_2), \bar{p}(x_1)\bar{p}(x_2)$ from $p(x_1, x_2), p(x_1)p(x_2)$, respectively [13–19].

Optimization of the nonparametric decision rule (5) with respect to the diffusion coefficients of the kernel functions c_1, c_2 can be simplified by setting in statistics (3) and (4) the values $c_v = c\bar{\sigma}_v, v = 1, 2$. Here, $\bar{\sigma}_v$ is the estimate of the standard deviation of the random variable x_v in the sample V . This statement is obvious, since a larger length of the interval of values x_v corresponds to a greater blur coefficient c_v of the kernel function $\Phi(u_v), v = 1, 2$. A similar approach was used in the construction of fast procedures for the optimization of nonparametric estimates of the kernel-type probability density [20–23].

The values of the estimates of the standard deviations $\bar{\sigma}_v$ are determined from the statistical data of the sample V :

$$\bar{\sigma}_v = \left(\frac{1}{n-1} \sum_{i=1}^n (x_v^i - \bar{x}_v)^2 \right)^{1/2}, \quad v = 1, 2.$$

Here, \bar{x}_v is the average value of the random variable x_v , which is calculated from the sample V .

Therefore, it becomes possible to optimize the nonparametric pattern recognition algorithm (5) using only one parameter c of the blur coefficients of the kernel functions.

Let us determine the estimates of the probabilities of pattern recognition errors $\bar{\rho}_1(\bar{c}_1, \bar{c}_2), \bar{\rho}_2(\bar{c}_1, \bar{c}_2)$ by the decision rule (5) based on the initial statistical data V for the optimal blur coefficients $\bar{c}(1) = (\bar{c}_1(1), \bar{c}_2(1)), \bar{c}(2) = (\bar{c}_1(2), \bar{c}_2(2))$ of the kernel functions of statistics $\bar{p}(x_1)\bar{p}(x_2), \bar{p}(x_1, x_2)$, respectively.

The values $\bar{\rho}_t(\bar{c}(1), \bar{c}(2))$ are calculated in the sliding exam mode for the sample V , assuming that its elements belong to the class Ω_t :

$$\bar{\rho}_t(\bar{c}(1), \bar{c}(2)) = \frac{1}{n} \sum_{j=1}^n 1(\delta(j), \bar{\delta}(j)), \quad t = 1, 2,$$

where $\delta(j) = t$ are indications of type $x^j = (x_1^j, x_2^j) \in \Omega_t$;

$$\bar{\delta}(j) = \begin{cases} t & \text{if } x^j \in \Omega_t; \\ 0 & \text{if } x^j \notin \Omega_t, \end{cases}$$

is the decision of algorithm (5) about the situation x^j belonging to one of the classes $\Omega_t, t = 1, 2$.

When calculating $\bar{\rho}_t(\bar{c}(1), \bar{c}(2))$ according to the method of the sliding exam method, the situation $x^j = (x_1^j, x_2^j)$ from the sample V , which is submitted for control to algorithm (5), is excluded from the process of generating statistics (3) and (4).

The indicator function is determined by the expression

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

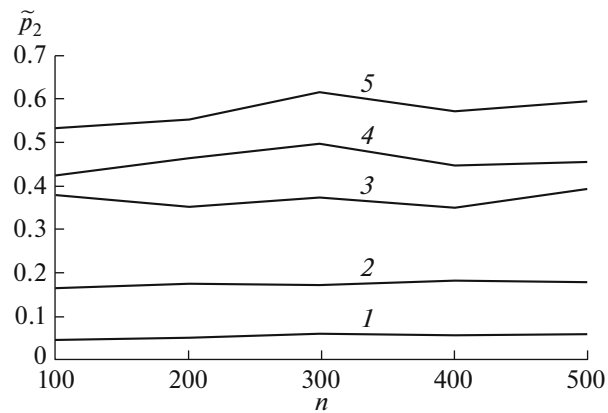


Fig. 1. Dependence of the averaged estimates of the probabilities of errors in belonging of elements of V to independent random variables x_1, x_2 on the sample size n (curves 1–5 correspond to the values of the correlation coefficients $r = 0.9, 0.7, 0.45, 0.33$, and 0).

Let us denote by $\bar{\rho}_t$ the minimum value of the estimate of the probability of a pattern recognition error under the assumption that the elements of the sample V belong to the class Ω_t , $t = 1, 2$. Compare the values $\bar{\rho}_1, \bar{\rho}_2$.

The hypothesis H_0 is valid if $\bar{\rho}_1 < \bar{\rho}_2$. Otherwise, for $\bar{\rho}_2 < \bar{\rho}_1$ the random variables x_1 and x_2 are dependent.

It is natural that with limited volumes n of the V sample, the problem of confidence estimation of the probabilities of image recognition errors arises. To solve it, we can use the traditional method of confidence estimation of probabilities [5] or the Kolmogorov–Smirnov criterion [24].

For example, when using the Kolmogorov–Smirnov criterion, the deviation $\bar{D}_{12} = |\bar{\rho}_1 - \bar{\rho}_2|$ is compared with the threshold value

$$D_\beta = \sqrt{-\ln(\beta/2)/n}.$$

Here, β is the probability (risk) to reject the hypothesis \bar{H}_0 : $\rho_1(c_1, c_2) = \rho_2(c_1, c_2)$. If the relation $\bar{D}_{12} < D_\beta$ is satisfied, then the hypothesis \bar{H}_0 is valid and the risk of rejecting it does not exceed the value β . If $\bar{D}_{12} > D_\beta$, the hypothesis \bar{H}_0 is rejected.

ANALYSIS OF THE RESULTS OF A COMPUTATIONAL EXPERIMENT

Let us investigate the dependence of the effectiveness of the proposed method for testing the hypothesis of the independence of two-dimensional random variables on the volume of the initial statistical data. We will assume that the random variables x_1 and x_2 have Gaussian distribution laws. When generating the values x_1, x_2 in the V sample, we use the random variable generators

$$x_1^i = M(x_1) + \sigma_1 \left(\sum_{j=1}^{12} \varepsilon_1^j - 6 \right), \quad x_2^i = M(x_2) + \sigma_2 \left(\sum_{j=1}^{12} \varepsilon_2^j - 6 \right), \quad i = \overline{1, n},$$

where $M(x_1)$ is the mathematical expectation of the random variable x_1 , σ_1 , and σ_2 are standard deviations of x_1 and x_2 , and ε_1 and ε_2 are random variables with uniform distribution laws over the interval $[0; 1]$.

When selecting the values σ_1, σ_2 , the dependence index (correlation coefficient) between the random variables x_1, x_2 changes, which is calculated from the obtained statistical data V . The values of $\bar{\rho}_1^t, \bar{\rho}_2^t$ for a specific volume $n(t) = n$ of the sample V are determined 50 times. The obtained data $\bar{\rho}_1^t, \bar{\rho}_2^t$, $t = \overline{1, 50}$ are averaged, and their results are denoted by $\tilde{\rho}_1, \tilde{\rho}_2$ and shown in Fig. 1.

The results of computational experiments confirm the effectiveness of the proposed method. With the correlation coefficient $r \geq 0.35$, the application of the method under consideration makes it possible to

exclude errors in assigning the initial statistical data V to independent random variables in 50 computational experiments. Let us denote this estimate of the probability of confirming the hypothesis H_0 by $\bar{P}_1 = 0$. If $r = 0$, the values are $\bar{P}_1 = 0.6, 0.62, \text{ and } 0.8$ with the volume of statistical data $n = 100, 200, \text{ and } 500$ respectively.

Let us analyze the values $\tilde{\rho}_1, \tilde{\rho}_2$, which determine the criterion for testing the hypothesis H_0 in computational experiments. With an increase in the correlation coefficient r , a decrease in the average estimate of the error probability $\tilde{\rho}_2$ of belonging of elements of the sample V to the class Ω_2 of values of dependent random variables is observed. For example, when increasing r in the interval $[0.45; 0.9]$ the estimate of the probability of a pattern recognition error $\tilde{\rho}_2$ decreases from 0.37 to 0.055. This fact is explained by a decrease in the area of intersection of the classes Ω_1, Ω_2 and, as a consequence, an increase in the values of the kernel estimate of the probability density $\bar{p}(x_1, x_2)$ in comparison with $\bar{p}(x_1)\bar{p}(x_2)$ in the nonparametric decision rule (5), which leads to a decrease in $\tilde{\rho}_2$. With a decrease in r in the range $[0.33; 0]$, estimates of the probabilities of the pattern recognition error $\tilde{\rho}_2$ increase from 0.45 to 0.53. Under these conditions, there is a tendency for the region of intersection of the classes Ω_1, Ω_2 and the values $\tilde{\rho}_2, \tilde{\rho}_1$ to converge as a criterion for the identity of the distribution laws $p(x_1, x_2), p(x_1)p(x_2)$ of the compared random variables.

The stability of the proposed method to the volume of initial statistical data was found at specific values of the correlation coefficient, which manifests itself in close values of $\tilde{\rho}_2$ at $n \in [100; 500]$. For example, for $r = 0.9$, the values are $\tilde{\rho}_2 \in [0.044; 0.06]$, and under the conditions $r = 0.45$ the values are $\tilde{\rho}_2 \in [0.35; 0.39]$ (see Fig. 1). The noted pattern weakens with decreasing values of r . This conclusion is confirmed by the values of $\tilde{\rho}_2$ in the interval $[0.423; 0.496]$ for $r = 0.33$.

The above statements are reliable, which is verified using the Kolmogorov–Smirnov criterion with the risk $\beta = 0.05$ to reject the hypothesis being tested. Doubtful decisions appear at values of r close to zero. Under these conditions, the proposed method provides a reliable solution for $n \geq 300$. For example, for $n = 300, 400, 500$ the values are $\bar{D}_{12} = 0.23, 0.141, \text{ and } 0.189$, that exceed the thresholds $D_\beta = 0.111, 0.096, \text{ and } 0.086$. The results confirm the hypothesis of the independence of random variables.

The results of computational experiments were compared with the confidence limits of the correlation coefficient

$$\tanh\left(\frac{1}{2} \ln \frac{1 + \bar{r}}{1 - \bar{r}} - \frac{\varepsilon_\alpha}{\sqrt{n-3}}\right) < r < \tanh\left(\frac{1}{2} \ln \frac{1 + \bar{r}}{1 - \bar{r}} + \frac{\varepsilon_\alpha}{\sqrt{n-3}}\right),$$

where ε_α is defined by the relation $2F(\varepsilon_\alpha) = \alpha$ and \bar{r} is the estimate of the correlation coefficient. Here, $F(\varepsilon_\alpha)$ is the Laplace function, α is confidence factor, and $\tanh(\cdot)$ is the hyperbolic tangent. For these conditions at $\bar{r} = 0, \alpha = 0.95$, and $\varepsilon_\alpha = 1.96$, the confidence limits for the correlation coefficient are determined by the intervals $r \in (\pm 0.196), (\pm 0.139), (\pm 0.113), (\pm 0.098), (\pm 0.088)$, which correspond to the volumes of statistical data $n = 100, 200, 300, 400, 500$.

Let us compare the effectiveness of the proposed method with the approach that uses the correlation coefficient as a criterion for the linear dependence between random variables. To do this, we determine the estimates of the probabilities of decisions with respect to the hypothesis H_0 , taken in accordance with the proposed method, under the conditions of the value of the correlation coefficient $r = 0.196$. Note that this value of r corresponds to its confidence boundary at $n = 100$ and $\alpha = 0.95$. Under these conditions, the estimate of the probability of confirming the hypothesis H_0 corresponds to the value of $\bar{P}_1 = 0.4$, and its refutation is $\bar{P}_2 = 0.6$ in 50 computational experiments. By the values of \bar{P}_1, \bar{P}_2 , the proposed method is more sensitive to changes in the indicator r of the linear dependence between the random variables x_1, x_2 . The results are consistent with the traditional approach of testing the hypothesis of linear dependence of random variables. However, the presented method applies to the conditions of nonlinear dependence between random variables.

CONCLUSIONS

The method proposed in this paper for testing the hypothesis of the independence of random variables bypasses the problem of decomposing the range of values of random variables into multidimensional intervals, which is characteristic of the Pearson criterion. To solve this problem, we use a nonparametric pattern recognition algorithm that meets the maximum likelihood criterion. Optimization of the kernel

probability density estimations of the blur coefficients is carried out from the condition of the maximum likelihood function. Under the assumption of independence or dependence of random variables in the initial statistical data, estimates of the probabilities of pattern recognition errors are determined. Based on their minimum value, a decision is made about the independence or dependence of random variables.

The effectiveness of the proposed method is confirmed by the results of computational experiments when testing the hypothesis of the independence of a two-dimensional random variable the components of which are characterized by normal distribution laws. It was found that with a correlation coefficient $r \geq 0.35$ between random variables, the proposed method accurately rejects the initial hypothesis with the volume of initial statistical data from 100 to 500. With independent random variables in conditions when the correlation coefficient is zero, the initial hypothesis is confirmed by probability estimates of 0.6, 0.62, and 0.8 with the volume of statistical data $n = 100, 200, \text{ and } 500$, respectively. The stability of the values of the used criterion for testing the hypothesis under consideration to changes in the volume of statistical data under specific experimental conditions is observed.

Promising research in this direction is the application of the proposed technique to test the hypothesis of a nonlinear relationship between random variables and the formation of a set of independent random variables, which will simplify the task of synthesizing effective information processing algorithms.

FUNDING

The research was carried out with the financial support of the Russian Foundation for Basic Research, the Government of the Krasnoyarsk krai, and the Krasnoyarsk Regional Science Foundation (project no. 20-41-240001).

REFERENCES

1. A. V. Lapko and V. A. Lapko, "Properties of nonparametric estimates of multidimensional probability density of independent random variables," *Inf. Sci. Control Syst.* **31** (1), 166–174 (2012).
2. A. V. Lapko and V. A. Lapko, "Nonparametric estimation of probability density of independent random variables," *Inf. Sci. Control Syst.* **29** (3), 118–124 (2011).
3. A. V. Lapko and V. A. Lapko, "Effect of a priori information about independence multidimensional random variables on the properties of their nonparametric density probability estimates," *Sist. Upr. Inf. Tekhnol.* **48** (2.1), 164–167 (2012).
4. A. V. Lapko and V. A. Lapko, "Properties of the nonparametric decision function with a priori information on independence of attributes of classified objects," *Optoelectron., Instrum. Data Process.* **48**, 416–422 (2012). <https://doi.org/10.3103/S8756699012040139>
5. V. S. Pugachev, *Theory of Probability and Mathematical Statistics* (Fizmatlit, Moscow, 2002).
6. A. V. Lapko and V. A. Lapko, "Nonparametric algorithms of pattern recognition in the problem of testing a statistical hypothesis on identity of two distribution laws of random variables," *Optoelectron., Instrum. Data Process.* **46**, 545–550 (2010). <https://doi.org/10.3103/S8756699011060069>
7. 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," *Optoelectron., Instrum. Data Process.* **48**, 37–41 (2012). <https://doi.org/10.3103/S8756699012010050>
8. A. V. Lapko and V. A. Lapko, "A technique for testing hypotheses for distributions of multidimensional spectral data using a nonparametric pattern recognition algorithm," *Comput. Optics* **43**, 238–244 (2019). <https://doi.org/10.18287/2412-6179-2019-43-2-238-244>
9. E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.* **33**, 1065–1076 (1962). <https://doi.org/10.1214/aoms/1177704472>
10. V. A. Epanechnikov, "Non-parametric estimation fo a multivariate probability density," *Theory Probab. Its Appl.* **14**, 153–158 (1969). <https://doi.org/10.1137/1114019>
11. R. P. W. Duin, "On the choice of smoothing parameters for parzen estimators of probability density functions," *IEEE Trans. Comput.* **C-25**, 1175–1179 (1976). <https://doi.org/10.1109/TC.1976.1674577>
12. Z. I. Botev and D. P. Kroese, "Non-asymptotic bandwidth selection for density estimation of discrete data," *Methodol. Comput. Appl. Probab.* **10**, 435 (2008). <https://doi.org/10.1007/s11009-007-9057-z>
13. M. Rudemo, "Empirical choice of histogram and kernel density estimators," *Scand. J. Stat.* **9**, 65–78 (1982).
14. A. W. Bowman, "A comparative study of some kernel-based non-parametric density estimators," *J. Stat. Comput. Simul.* **21**, 313–327 (1982). <https://doi.org/10.1080/00949658508810822>

15. P. Hall, "Large-sample optimality of least squares cross-validation in density estimation," *Ann. Statist.* **11**, 1156–1174 (1983).
16. M. Jiang and S. B. Provost, "A hybrid bandwidth selection methodology for kernel density estimation," *J. Stat. Comput. Simul.* **84**, 614–627 (2014). <https://doi.org/10.1080/00949655.2012.721366>
17. S. Dutta, "Cross-validation revisited," *Commun. Stat. Simul. Comput.* **45**, 472–490 (2016). <https://doi.org/10.1080/03610918.2013.862275>
18. N.-B. Heidenreich, A. Schindler, and S. Sperlich, "Bandwidth selection for kernel density estimation: a review of fully automatic selectors," *ASTA Adv. Stat. Anal.* **97**, 403–433 (2013). <https://doi.org/10.1007/s10182-013-0216-y>
19. Q. Li and J. S. Racine, *Nonparametric Econometrics: Theory and Practice* (Princeton Univ. Press, Princeton, 2007).
20. A. V. Lapko and V. A. Lapko, "Method of fast bandwidth selection in a nonparametric classifier corresponding to the a posteriori probability maximum criterion," *Optoelectron., Instrum. Data Process.* **55**, 597–605 (2019). <https://doi.org/10.3103/S8756699019060104>
21. D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, New Jersey, 2015). <https://doi.org/10.1002/9780470316849>
22. S. J. Sheather, "Density estimation," *Stat. Sci.* **19**, 588–597 (2004). <https://doi.org/10.1214/088342304000000297>
23. B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).
24. A. S. Sharakshane, I. G. Zheleznov, and V. A. Ivinskii, *Complex Systems* (Vysshaya Shkola, Moscow, 1977).

Translated by T. N. Sokolova