Verification of recovery of distribution density function of the random variable by methods of nonparametric statistics

V N Syzrantsev¹, K V Syzrantseva¹

¹Department of Machines and equipment of oil and gas industry, Industrial University of Tyumen, 38, Volodarskogo str., Tyumen, 625000, Russia
²Department of Cybernetic systems, Industrial University of Tyumen, 38, Volodarskogo str., Tyumen, 625000, Russia

E-mail: kv.syzr@gmail.com

Abstract. The paper considers an algorithm of recovery of distribution density function of a random variable by methods of nonparametric statistics. For samples of various length (from 20 to 2000) of normal distribution the quantile estimates of a random variable are calculated by methods of parametrical and nonparametric statistics. It is established that errors of the quantile estimates obtained by parametrical and nonparametric methods are values of the same order.

1. Introduction

When assessing strength reliability of elements of different machines, stresses and their limit values obtained during an experiment or calculation in the form of random variable $x_i, i = 1, N$ samples are used. Consequently, determining failure-free operation probability of the elements and metal construction of machines requires knowledge of the functions of the above mentioned random variables distribution density.

In the practice of processing experimental data for solving strength reliability problems it is traditionally accepted [1, 2] that required function $P(X)$ is known up to the finite number of the parameters. In this case, the problem of density recovery is well-posed and its solution uses methods of parametric statistics [2, 3, 4, 5]. However, in general case, the class of functions to which $P(X)$ can belong is substantially wider than those discussed in papers [1, 2, 3, 4, 5] and many others.

To date, in the theories of probability and mathematical statistics over a hundred of random variables distribution laws have been studied. At the same time, the experience of processing experimental data on machine parts and metal constructions loading and durability during operation indicates that the distribution of random variables of operating and limit stresses, the number of cycles prior to parts and carrying systems destruction do not follow the standard laws described in the theory of parametric statistics.

To study the processes containing random variables not described by standard distributions nonparametric methods have been developed in the theory of mathematical statistics [6, 7, 8, 9, 10, 11, 12, 13]. When processing experimental data in problems of recovery of unknown function of random variables distribution density the methods of nonparametric statistics [14, 15, 16, 17, 18] based on Parzen-Rosenblatt estimates are mostly used [11, 12], or the decomposition method with respect to basic functions [6].
2. Nonparametric Estimates of the Distribution Density Function
Based on the physical meaning of $X$, we present the only requirement – its continuity and consider the problem of $P(X)$ recovery based on the existing sample $x_i, i = 1, N$ by means of nonparametric statistics [7, 8].

It is known [2, 4, 5, 8], that the probability distribution density is related to the probability distribution function by the ratio:

$$\int_{-\infty}^{y} P(x) dx = F(y),$$

presented in the form of:

$$\int_{-\infty}^{\theta(y-x)} P(x) dx = F(y), \quad \theta(s) = \begin{cases} 1, & \text{if } s \geq 0 \\ 0, & \text{if } s < 0 \end{cases},$$

where $\theta(s)$ is a function of the unit step (Heaviside function).

The optimal nonparametric estimate at each point $y$ for theoretical distribution function $F(y)$ is [8] empirical cumulative probability distribution function $F_N(y)$. If value $y$ exceeds $k$ of the sampling elements $x_i, i = 1, N$ by volume $N$ function $F_N(y)$ looks like:

$$F_N(y) = \frac{1}{N} \sum_{i=1}^{N} \theta(y-x_i).$$

With the growth of sample volume $N$ function $F_N(y)$ with probability of one uniformly approaches $F(y)$:

$$\Pr \left\{ \lim_{N \to \infty} \sup \left| F_N(y) - F(y) \right| = 0 \right\} = 1.$$  (4)

In statistical processing the data of experimental stress studies the right part of equation (2) – distribution function $F(y)$ is replaced by empirical distribution function $F_N(y)$ obtained on the basis of a sample of limited volume stresses. Therefore, the solution of the equation (2) will always be approximate. To recover the distribution density function by solving the equation (2) within the theory of nonparametric statistics, special procedures have been developed [6, 7, 8], providing equation solutions sequence convergence (2) to the required probability density $P(X)$ with increasing $N$ and taking into account the incorrect formulation of the problem (2), associated with the requirement to differentiate the inexact right part of the equation (2). Based on the physical meaning of the distribution density function, it has only positive values on the entire axis of change $X$ – from $-\infty$ to $+\infty$ – and fulfills the following condition:

$$\int_{-\infty}^{+\infty} P(x) dx = 1.$$  (5)

Under the condition of continuity of function $P(X)$ the solution (2) is the only one.

3. Algorithms for recovery of unknown distribution density function
The methods of distribution density function estimation proposed by Parzen and Rosenblatt [11, 12], use a smoothed empirical distribution function in the form of:
\[ F_N(y) = \frac{1}{N} \sum_{i=1}^{N} G \left( \frac{y-x_i}{h_N} \right), \]  

(6)

where \( G(t) \) is a steadily non-decreasing function from 0 to 1 of its argument; in this case \( G(t) = 1 - G(-t) \), i.e. \( G(t) \) is a symmetric function relative to zero; \( h_N \) is a fuzziness parameter (bandwidth).

After differentiation (6) we have:

\[ P_N(y) = F'_N(y) = \frac{1}{N \cdot h_N} \sum_{i=1}^{N} G' \left( \frac{y-x_i}{h_N} \right) = \frac{1}{N \cdot h_N} \sum_{i=1}^{N} K \left( \frac{y-x_i}{h} \right), \]  

(7)

where \( K(t) = G'(t) \) is distribution density \( G(t) \) or kernel function (kernel).

The theoretical studies \[8\] of the function (6) show that displacement and variation of the estimate (7) depend on kernel type \( K(t) \) and the value of fuzziness parameter \( h_N \). Various dependencies are used as kernel functions \[8, 9, 17\]. The most common of them, used in processing experiments data by assessing loading and durability of mechanical engineering products, are presented in Table 1.

Recovery of distribution density function by means of the Parzen–Rosenblatt method based on (7) is performed in two stages. At the first stage, kernel function \( K(t) \) is selected from among the known ones (Table 1 or others). At the second stage, the problem of determining an optimal value of fuzzing parameter \( h_N \) is solved.

In physical meaning, use of function (7) allows describing a random variable, the discrete values of which are recorded in the course of experimental studies, with continuous function which provides a possibility of its application to solve numerous practical problems of differential and integral methods.

To select the best one from among a finite number of functions \( K(t) \), it is required to have a selection criterion. Information functional of the form \[8, 13, 15\]:

\[ J = \int \ln [K(t)] P(t) dt = \int \ln [K(t)] dF(t), \]  

(8)

the maximum value of which fulfills condition \( K(t) = P(t) \) can be adopted as such a criterion.

Then the search for optimal \( h_N^* \) and \( K(t) \in K = \{K_1(t), \ldots, K_2(t)\} \) comes down to solving the following problem \[8, 16, 17\]:

\[ (h_N^*, K^*(t)) = \arg \max_{h_N, K(t)} J_N(h_N, K(t)) = \arg \max_{h_N, K(t)} \left[ \frac{1}{N} \sum_{i=1}^{N} \ln \left( \frac{1}{(N-1)h_N} \sum_{j=1}^{N-1} K \left( \frac{x_i - x_j}{h_N} \right) \right) \right], \]  

(9)

Determination of optimal value \( h_N^* \), as shown in papers \[6, 8\], is a more difficult task than a recovery of distribution density, since optimal value \( h_N^* \) depends on unknown distribution density, especially, its unknown derivatives. In practical applications density estimates are often required in particular areas; for example, in prediction of resource and reliability, first of all, density estimate in the distribution tail area is very important. Therefore, in solving this problem, the algorithms providing determination of optimal values of parameter \( h_N^* \) on the basis of only available sample \( x_i, i = 1, N \) of random variable \( X \) are required.

For some functions \( K(t) \) in Table 1, the dependences for calculating estimates of optimal value \( h_N^* \) based on different estimates of sample \( x_i, i = 1, N \) have been obtained. For example, when using kernel function as normal distribution (Table 1):
\[ K\left( \frac{y-x_i}{h_N} \right) = \frac{1}{\sqrt{2\cdot\pi}} \exp \left[ -0.5 \left( \frac{y-x_i}{h_N} \right)^2 \right]. \] (10)

**Table 1.** Kernel functions [8, 9, 17].

| Kernel function                  | Uniform kernel | Triangle kernel No.1 | Triangle kernel No.2 |
|----------------------------------|----------------|----------------------|----------------------|
| \[ K_1(t) = \begin{cases} 0.5|t| & \text{if } |t| \leq 1 \\ 0 & \text{if } |t| > 1 \end{cases} \] |                |                      |                      |
| \[ K_2(t) = \begin{cases} 1-|t| & \text{if } |t| \leq 1 \\ 0 & \text{if } |t| > 1 \end{cases} \] |                |                      |                      |
| \[ K_3(t) = \begin{cases} \frac{1}{\sqrt{6}} - \frac{|t|}{6} & \text{if } |t| \leq \sqrt{6} \\ 0 & \text{if } |t| > \sqrt{6} \end{cases} \] |                |                      |                      |

| Quadratic kernel                 | Parabolic kernel | Epanechnikov kernel |
|----------------------------------|------------------|---------------------|
| \[ K_4(t) = \frac{3\cdot(1-t^2)}{4} | t | \leq 1 \] | \[ K_5(t) = \frac{9}{8} \left( 1 - \frac{5}{3} |t| \right) | | 0 \leq |t| \leq 1 \] | \[ K_6(t) = \frac{3}{4\sqrt{5}} \left( 1 - \frac{2}{5} |t| \right) | | 0 \leq |t| \leq \sqrt{5} \] |

| Gaussian kernel                  | Laplace kernel   | Fisher kernel       |
|----------------------------------|------------------|---------------------|
| \[ K_7(t) = \frac{1}{\sqrt{2\cdot\pi}} \exp \left( -\frac{t^2}{2} \right) \] | \[ K_8(t) = \frac{1}{2} \exp(-|t|) \] | \[ K_9(t) = \frac{1}{2\cdot\pi} \left( \frac{2}{\sin \left( \frac{t}{2} \right)} \right) \] |

| Logistics kernel                 | Trigonometric kernel | Cauchy kernel |
|----------------------------------|----------------------|---------------|
| \[ K_{10}(t) = \frac{\exp(-t)}{\left[ 1 + \exp(-t) \right]^2} \] | \[ K_{11}(t) = \left\{ \begin{array}{ll} 0.54 + 0.46 \cos(\pi \cdot t) & |t| \leq 1 \\ 0 & |t| > 1 \end{array} \right. \] | \[ K_{12}(t) = \frac{1}{\pi} \left( \frac{1}{1 + t^2} \right) \] |

| Vallee-Poussin                   | Jacobi kernel      | Chebyshev-Hermite kernel |
|----------------------------------|-------------------|-------------------------|
| \[ K_{13}(t) = \frac{1}{2\cdot\pi} \left( \frac{2}{t} \sin \left( \frac{t}{2} \right) \right)^2 \] | \[ K_{14}(t) = \left\{ \begin{array}{ll} \frac{15}{2} \left( 3 - 10 t^2 + 7 t^4 \right) & |t| \leq 1 \\ 0 & |t| > 1 \end{array} \right. \] | \[ K_{15}(t) = \frac{3}{2} \left( 1 - \frac{t^2}{3} \right) \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{t^2}{2} \right) \] |

Optimal value of parameter \( h_N^* \) («bandwidth»), is defined by formula [6]:

\[ h_N^* = D_N \cdot N^{-0.2}, \] (11)

where \( D_N \) is sampling variance, calculated based on the available values sample \( x_i, i = 1, N \):

\[ D_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left( x_i - \frac{1}{N} \sum_{i=1}^{N} x_i \right)^2. \] (12)

As a result, to estimate the density with kernel (10) and fuzziness parameter (11) based on (7) we have the following expression:

\[ P_N(y) = \frac{1}{\sqrt{2\cdot\pi \cdot N \cdot h_N^*}} \sum_{i=1}^{N} \exp \left[ -0.5 \left( \frac{y-x_i}{h_N^*} \right)^2 \right]. \] (13)
Analyzing (13), it is easy to see that the implementation of the Parzen-Rosenblatt method involves two steps of calculation. First, rough characterization of the sample, sampling variance (12), is calculated, which is further used to correct the estimation of distribution density (13) via fuzziness parameter $h_N^*$. Since value $D_N$ is sensitive to emissions and does not reflect the nature of the density function change (unimodal, multimodal) the information about the distribution density obtained with $D_N$ may not be sufficient to solve the problem correctly with the method under consideration.

To solve the problem of recovery of distribution density unknown function using (9) on the basis of the kernel functions presented in Table 1, a set of programs in the Mathcad system has been developed. The search for function maximum (9) for each kernel function $K(t)$ from Table 1 by value $h_N$ is carried out with the numerical method taking into account the initial value $h_N$ calculated by expression (11).

4. Task of Verifying the Recovery of Unknown Random-Variable Distribution Density Function by Means of Nonparametric Statistical Methods

To date, the apparatus of nonparametric statistics has been widely used in processing experimental data to solve the image (target) identification tasks, as well as in medicine, cryptography, macroeconomics, reliability screening for parts and metal structures in various machines and equipment. At the same time, nonparametric statistical methods are reluctantly applied for processing experimental data, which is explained by wider confidence intervals of a random variable, provided by nonparametric statistics in comparison with traditional methods. In this paper, using the samples of a random variable of a different length (from 20 to 2000) from the normal distribution, we calculated its quantile estimators through parametric and nonparametric statistical methods and analyzed the error of these estimators to verify the recovery of the density function by means of nonparametric statistical methods.

It is known that the normal distribution density function looks like:

$$P(x) = \frac{1}{s\sqrt{2\cdot \pi}} \exp \left[-0.5\left(\frac{x-m}{s}\right)^2\right], \quad (14)$$

where $m$ and $s$ constitute an average deviation and a mean square deviation of the random variable $x$, respectively.

Let us assume that $x$ changes within the limits of $x_{\min} = 10$ and $x_{\max} = 2000$, and the distribution parameters (14) have the values of $m=800$; $s=80$.

The quantile values ($x_{\alpha}$) of the random variable $x$ are determined by solving the following equation through the calculus of approximations:

$$\int_{x_{\min}}^{x_{\alpha}} P(x)dx = \frac{1}{s\sqrt{2\cdot \pi}} \int_{x_{\min}}^{x_{\alpha}} \exp \left[-0.5\left(\frac{x-m}{s}\right)^2\right]dx = \alpha. \quad (15)$$

The calculation results $x_{\alpha}$ based on the equation (15) with $m=800$, $s=80$ and $x_{\min} = 10$ for the range of values $\alpha$, which correspond to the «tails» of the normal distribution, are given at the top of Tables 2 and 3.

| $\alpha$ | 0.005 | 0.010 | 0.050 | 0.100 |
|----------|-------|-------|-------|-------|
| Normal distribution low (14) |       |       |       |       |
| Value $x_{\alpha}$ | 593.933 | 613.901 | 668.410 | 697.475 |
| $n = 20$ |       |       |       |       |
|   | \( \hat{x}_a \)   | \( x_\alpha \) | \( x_\alpha^{\hat{\cdot}} \) | \( x_\alpha^{\hat{\cdot}} \) |
|---|---|---|---|---|
| **Value** | 642.191 | 659.440 | 706.509 | 731.613 |
| **Error (%)** | +8.12 | +7.42 | +5.70 | +4.89 |
| **Value** | 626.205 | 642.490 | 690.990 | 718.037 |
| **Error (%)** | +5.43 | +4.66 | +3.38 | +2.95 |
| **n = 50** | | | | |
| **Value** | 594.826 | 615.220 | 670.905 | 700.593 |
| **Error (%)** | +8.12 | +7.42 | +5.70 | +4.89 |
| **Value** | 533.821 | 560.640 | 643.170 | 683.530 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |
| **n = 100** | | | | |
| **Value** | 587.842 | 609.090 | 667.100 | 698.025 |
| **Error (%)** | +5.43 | +4.66 | +3.38 | +2.95 |
| **Value** | 543.635 | 564.880 | 639.260 | 681.840 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |
| **n = 200** | | | | |
| **Value** | 597.553 | 613.800 | 671.530 | 700.400 |
| **Error (%)** | +5.43 | +4.66 | +3.38 | +2.95 |
| **Value** | 568.535 | 590.500 | 659.545 | 697.330 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |
| **n = 500** | | | | |
| **Value** | 564.738 | 618.160 | 672.185 | 700.990 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |
| **Value** | 588.884 | 607.320 | 666.440 | 697.980 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |
| **n = 2000** | | | | |
| **Value** | 597.553 | 613.800 | 671.530 | 700.400 |
| **Error (%)** | +5.43 | +4.66 | +3.38 | +2.95 |
| **Value** | 568.535 | 590.500 | 659.545 | 697.330 |
| **Error (%)** | -10.12 | -8.67 | -3.77 | -1.99 |

**Table 3.** The results of calculation of \( x_\alpha \) (right tail of distribution).

|   | \( x_\alpha \) | \( x_\alpha^{\hat{\cdot}} \) | \( x_\alpha^{\hat{\cdot}} \) | \( x_\alpha^{\hat{\cdot}} \) |
|---|---|---|---|---|
| **Value** | 902.525 | 931.590 | 986.100 | 1006.09 |
| **n = 20** | | | | |
| **Value** | 908.696 | 933.800 | 980.883 | 998.140 |
| **Error (%)** | +0.68 | +0.24 | -0.53 | -0.79 |
| **Value** | 927.680 | 952.457 | 993.856 | 1007.95 |
| **Error (%)** | -2.78 | +2.23 | +0.78 | +0.18 |
| **n = 50** | | | | |
Using the random number generator (the algorithm and the program to realize it are presented in
the work [16]), according to the law (14), we generate a sample with the length of $n = 2000$ for the
random variable $x_i, \ i = 1, n$, which becomes the basis for determining the parameters of the law (14),
with the sample volume sequentially increasing:

$$\hat{m} = \frac{1}{n} \sum_{i=1}^{n} x_i, \ \hat{s} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{m})^2}, \ n = 20, 50, 100, 200, 500, 2000.$$ (16)

In this case, the normal distribution density function looks like the following:

$$\hat{P}(x) = \frac{1}{\hat{s}\sqrt{2 \cdot \pi}} \exp \left[-0.5 \left( \frac{x - \hat{m}}{\hat{s}} \right)^2 \right],$$ (17)

and the quantile estimators $\hat{x}_\alpha$ of the random variable $x$, calculated on the basis of the sample, are
determined by solving an equation similar to (15):

$$\int_{x_{min}}^{\hat{x}_\alpha} \hat{P}(x) dx = \frac{1}{\hat{s}\sqrt{2 \cdot \pi}} \int_{x_{min}}^{\hat{x}_\alpha} \exp \left[-0.5 \left( \frac{x - \hat{m}}{\hat{s}} \right)^2 \right] dx = \alpha.$$ (18)
As the truth values \( x_\alpha \) are known, we are able to analyze the errors \( \delta \hat{x}_\alpha \) of the obtained estimators \( \hat{x}_\alpha \):

\[
\delta \hat{x}_\alpha = 100 \left( \frac{\hat{x}_\alpha - x_\alpha}{x_\alpha} \right).
\] (19)

The values \( \hat{x}_\alpha, \delta \hat{x}_\alpha \) calculated with different sample volumes of the random variable \( x \) \( (n = 20, 50, 100, 200, 500, 2000) \) are presented in Tables 2 and 3.

In Figures 1–6, the dashed line shows the density function \( P(x) \) and histograms of the random variable \( x \), which correspond to different sample volumes.

\[\text{Figure 1. Distribution density functions } P(x), P_n(x) \text{ and the histogram of distribution of a random variable } x \text{ at } n = 20\]

\[\text{Figure 2. Distribution density functions } P(x), P_n(x) \text{ and the histogram of distribution of a random variable } x \text{ at } n = 50\]
Figure 3. Distribution density functions $P(x)$, $P_n(x)$ and the histogram of distribution of a random variable $x$ at $n = 100$

Figure 4. Distribution density functions $P(x)$, $P_n(x)$ and the histogram of distribution of a random variable $x$ at $n = 200$
Next, having the samples of the random variable $x$, we recover the unknown distribution density function $P_n(x)$ by means of nonparametric statistical methods. Using the Gaussian as the kernel function (Table 1), we present the desired function $P_n(x)$ as follows:

$$
P_n(x) = \frac{1}{\sqrt{2\pi n h}} \sum_{i=1}^{n} \exp \left[ -0.5 \left( \frac{x-x_i}{h_n^*} \right)^2 \right],
$$

where the optimal value of fuzziness parameter $h_n^*$ corresponds to the functional maximum:
When solving the problem (21), the initial estimate $h_n$ is calculated according to the dependence (11), taking into account (12).

Using the distribution density function $P_n(x)$, the quantile estimators $\hat{x}_a$ of the random variable $x$ are determined by solving the following equation through the calculus of approximations:

$$\int_{x_{\text{min}}}^{\hat{x}_a} P_n(x)dx = \frac{1}{n \cdot h_n^2 \sqrt{2 \cdot \pi}} \frac{\hat{x}_a}{\sum_{i=1}^{n} \exp \left[-0,5 \left( \frac{x_i - x_{\text{min}}}{h_n} \right)^2 \right]} dx = \alpha.$$  \hspace{1cm} (22)

With the known truth values $x_a$, the errors of the estimators $\tilde{x}_a$ are calculated according to the similar formula (19) of dependence:

$$\delta \tilde{x}_a = 100 \left( \frac{\tilde{x}_a - x_a}{x_a} \right).$$  \hspace{1cm} (23)

The values $\hat{x}_a$, $\tilde{x}_a$, $\delta \hat{x}_a$, obtained with different sample volumes of the random variable $x$ ($n = 20, 50, 100, 200, 500, 2000$) are given in Tables 2 and 3, and in Figures 1–6 the function $P_n(x)$ recovered by means of nonparametric statistical methods with different $n$ is indicated with a solid line.

The analysis of the results presented in Tables 2 and 3 shows that for the left distribution «tail» the maximum error $|\delta \hat{x}_a| = 8.12\%$ (for the sample $n = 20$ and $\alpha = 0.005$), and $|\delta \tilde{x}_a| = 10.12\%$ (for the sample $n = 50$ and $\alpha = 0.005$). For the right distribution «tail» the maximum error $|\delta \hat{x}_a| = 3.41\%$ (for the sample $n = 2000$ and $\alpha = 0.995$), and $|\delta \tilde{x}_a| = 2.78\%$ (for the sample $n = 20$ and $\alpha = 0.900$). In other words, the errors of the quantile estimators obtained by means of parametric and nonparametric statistical methods constitute equal-order values.

5. Conclusion
In connection to the task of verifying the recovery of random-variable distribution density function having the normal distribution law by means of nonparametric statistical methods through correlating the values of the random variable quantile estimators in the left and right distribution «tails», the results presented in this work prove that the errors of these estimators calculated by means of parametric and nonparametric statistical methods with the sample length range from 20 to 2000 constitute equal-order values.

This work was supported by grant (project N9.6355.2017/BCh) of government order of Ministry of education and science of Russian Federation for the period 2017–2019 in Tyumen Industrial University.

References
[1] Kapur K C and Lamberson L R 1977 Reliability in Engineering Design (New York: John Wiley & Sons) 586 p
[2] Gnedenko B V 1962 The Theory of Probability Physics Today 15 iss. 10 pp 62–64
[3] Gadolina I and Zaynetdinov R 2018 The estimation of the sufficient random loading realization length in the problem of machine parts longevity Proc. IEEE 9th Int. Conf. on Dependable Systems, Services and Technologies (Kiev) pp 159–162
[4] Collins J A 1981 Failure of Materials in Mechanical Design: Analysis, Prediction, Prevention (New York: John Wiley & Sons) 630 p
[5] Timashev S A, Bushinskaya A V, Malukova M G and Poluyan L V 2013 Integrity and Safety of Pipeline Systems ed S A Timashev (Yekaterinburg: UB RAS) 590 p
[6] Vapnik V 1999 An overview of statistical learning theory IEEE Transactions on Neural Networks 10 iss. 5 pp 988–999
[7] Orlov A I 2006 Applied statistics (Moscow: Exam) 671 p (in Russian)
[8] Simakhin V A 2006 Nonparametric robust regression estimate Proc. SPIE 13th Joint Int. Symp. on Atmospheric and Ocean Optics vol.6522 (Tomsk) pp 130–139
[9] Koshkin G M 2016 Smooth nonparametric estimation of the failure rate function and its first two derivatives Russian Physics Journal 59 iss. 6 pp 833–844
[10] Porshnev A S and Koposov A S 2014 Quantitative estimation peculiarities of densities blur parameter of random sequences distribution in Parzen-Rosenblatt approximation In the World of Scientific Discoveries, Series B 2 №1 2014 pp 120–129
[11] Parzen E 1962 On estimation of a probability density function and mode Ann. Math. Statist. 33 No 3 pp 1065–1076
[12] Rozenblatt M 1956 Remarks on some nonparametric estimates of a density functions Ann. Math. Statist. 27 No 3 pp 832–837
[13] Timashev S A and Tyrsin A N 2011 Entropy approach to risk-analysis of critical infrastructures systems Proc. First Int. Conf. on Vulnerability and Risk Analysis and Management and Fourth Int. Symp (Hyattsville) pp 147–154
[14] Syzrantseva K and Syzrantsev V 2017 Determination of parameters of endurance limit distribution law of material by the methods of nonparametric statistics and kinetic theory of high-cycle fatigue Key Engineering Materials 736 pp 52–57
[15] Syzrantseva K V 2009 Predicting the failure of gear transmissions by nonparametric statistics Russian Engineering Research 29 No 2 pp 1206–1208
[16] Syzrantsev V N, Golofast S L and Nevelev Ya P Calculation of equipment durability based on the methods of the distribution-free statistics (Novosibirsk: Nauka) 216 p (in Russian)
[17] Syzrantsev V N, Novoselov V V, Sozonov P M and Golofast S L Estimation of safety and durability of technological pipelines by methods of nonparametric statistics (Novosibirsk: Nauka) 172 p (in Russian)
[18] Syzrantseva K and Syzrantsev V 2016 Reliability estimation of machine parts with complicated geometry on a base of methods of nonparametric statistics Journal of Engineering and Applied Sciences 11 iss. 2 pp 204–209