

# Modification of nonparametric probability density estimation of Parzen's type with implicitly given form of kernel function

Alexandr V. Lapko  
Institute of Computer Modelling SB RAS  
Reshetnev Siberian State Aerospace University  
Krasnoyarsk, Russia  
e-mail: lapko@icm.krasn.ru

Vasily A. Lapko  
Institute of Computer Modelling SB RAS  
Reshetnev Siberian State Aerospace University  
Krasnoyarsk, Russia  
e-mail: valapko@yandex.ru

Ekaterina A. Yuronen  
Reshetnev Siberian State Aerospace University  
Siberian Federal University  
Krasnoyarsk, Russia  
e-mail: lapko@icm.krasn.ru

**Abstract** – The new nonparametric probability density estimation based on use of the smoothing operator is offered and investigated. It has smaller dispersion in comparison with probability density estimation like Rosenblatt-Parzen.

**Keywords** – probability density, nonparametric estimation Rosenblatt-Parzen, approximation properties, kernel function, Sturges' rule, Heinhold-Gaede rule

For on selection of  $x^i, i = \overline{1, n}$  statistically independent observations of a random value  $x$  we will use its nonparametric assessment [1]

$$\bar{p}(x) = \frac{1}{nc} \sum_{i=1}^n \Phi\left(\frac{x-x^i}{c}\right). \quad (1)$$

The kernel function  $\Phi(u)$  obeys

$$\Phi(u) = \Phi(-u), \quad 0 \leq \Phi(u) < \infty,$$

$$\int \Phi(u) du = 1, \quad \int u^2 \Phi(u) du = 1,$$

$$\int u^m \Phi(u) du < \infty, \quad 0 \leq m < \infty,$$

and coefficients of a diffuseness  $c = c(n)$  of function  $\Phi(u)$  such that  $c(n) \rightarrow 0$  at  $n \rightarrow \infty$ . The infinite limits of integration fall in this case and further.

Nonparametric probability density estimation in the form of (1) is generalized for a many-dimensional case in works [2, 3]. Conditions of its asymptotic unbiasedness and solvency are established. Nonparametric probability density estimates which are based on the procedure of compression of input statistical data are developed for conditions of selections of large volume [4]. The methods of optimum sampling of a range of values of random values [5] and confidence estimation of their probability density are offered on this basis [6].

At the restricted volumes of  $n$  of statistical data nonparametric assessment of a probability density of type (1) is characterized by the increased value of dispersion. For decrease of dispersion of traditional nonparametric probability density estimates the smoothing operators are used [7]. In this work the offered approach develops on use at creation of nonparametric assessment of a probability density with implicitly given form of kernel function of an optimum kernel of V. A. Epanechnikov. Properties of the received probability density assessment depending on conditions of their application are investigated.

## I. SYNTHESIS OF NONPARAMETRIC PROBABILITY DENSITY ESTIMATION

As an approximation on empirical data of a one-dimensional probability density  $p(x)$  we will accept statistics

$$\bar{p}_1(x) = \frac{1}{2\beta} \int_{x-\beta}^{x+\beta} \bar{p}(u) du, \quad (2)$$

where  $\bar{p}(u)$  - nonparametric assessment of a probability density (1).

Let's substitute in (2) expression, defining  $\bar{p}(u)$ , we will receive получим

$$\bar{p}_1(x) = \frac{1}{2\beta n c} \sum_{i=1}^n \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^i}{c}\right) du. \quad (3)$$

The statistics (3) falls into to a class of nonparametric estimates of a probability density with implicitly given kernel function

$$\frac{1}{2\beta c} \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^i}{c}\right) du. \quad (4)$$

The choice of values  $\beta$  and  $c$  by optimization of nonparametric assessment of a probability density (3) allows changing not only definition range of kernel function, but also its look. For example, at kernel function

$$\frac{1}{c} \Phi\left(\frac{u-x^i}{c}\right) = \begin{cases} 1/(2c), & \text{if } |u-x^i| \leq c, \\ 0, & \text{if } |u-x^i| > c \end{cases} \quad (5)$$

implicitly the given kernel (4) at  $c > \beta$  will take a form [7]

$$\begin{cases} 0, & \text{if } |x-x^i| \geq c + \beta \\ \frac{c + \beta - |x-x^i|}{4\beta c}, & \text{if } c - \beta < |x-x^i| < c + \beta \\ 1/(2c), & \text{if } |x-x^i| < c - \beta. \end{cases} \quad (6)$$

Let's accept as function (5) an optimum kernel of V. A. Epanechnikov [3]

$$\frac{1}{c} \Phi_0\left(\frac{x-x^i}{c}\right) = \begin{cases} \frac{3(5c^2 - (x-x^i)^2)}{20\sqrt{5}c^3}, & \text{if } |x-x^i| < \sqrt{5}c \\ 0, & \text{if } |x-x^i| \geq \sqrt{5}c. \end{cases} \quad (7)$$

Then implicitly the given kernel function (4) at  $\sqrt{5}c \geq \beta$  will take a form

$$\begin{cases} 0, & \text{if } |x-x^i| \geq \sqrt{5}c + \beta, \\ \alpha \left[ -5c^2 \left( |x-x^i| - \sqrt{5}c - \beta \right) - \frac{5\sqrt{5}c^3 - \left( |x-x^i| - \beta \right)^3}{3} \right], & \text{if } \sqrt{5}c - \beta \leq |x-x^i| < \sqrt{5}c + \beta, \\ \alpha \left[ 10c^2\beta - \frac{\left( |x-x^i| + \beta \right)^3 - \left( |x-x^i| - \beta \right)^3}{3} \right], & \text{if } |x-x^i| < \sqrt{5}c - \beta, \end{cases} \quad (8)$$

where  $\alpha = 3/(40\sqrt{5}c^3\beta)$ .

If the ratio is fair  $\sqrt{5}c < \beta$ , then function (4) looks as

$$\begin{cases} 0, & \text{if } |x-x^i| \geq \sqrt{5}c + \beta, \\ \alpha \left[ -5c^2 \left( |x-x^i| - \sqrt{5}c - \beta \right) - \frac{5\sqrt{5}c^3 - \left( |x-x^i| - \beta \right)^3}{3} \right], & \text{if } \beta - \sqrt{5}c \leq |x-x^i| < \sqrt{5}c + \beta, \\ 1/(2\beta), & \text{if } |x-x^i| < \beta - \sqrt{5}c. \end{cases} \quad (9)$$

The graphic illustration of a ratio (6) is presented on the figure 1a.

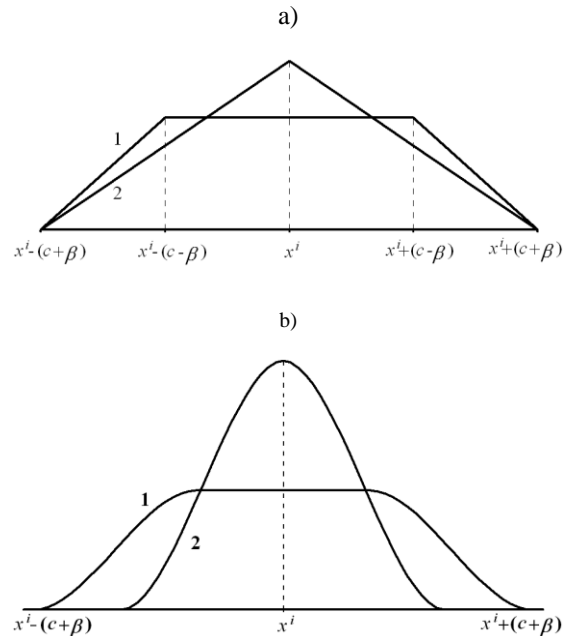


Fig. 1. A type of kernel function (5) in the conditions of (6) (a) at  $c > \beta$  (1),  $c = \beta$  (2) and conditions (8) (b) at  $\sqrt{5}c > \beta$  (1),  $\sqrt{5}c = \beta$  (2).

The received kernel functions (4) have property of a normalization.

Optimization of statistics (3) are carried out by the serial choice of parameters  $c$  and  $\beta$ . Taking into account a technique of synthesis of the studied nonparametric estimates of a probability density  $\bar{p}_1(x)$ ,  $\bar{p}_2(x)$  and  $\bar{p}(x)$  value of their coefficients of a diffuseness  $c$  of kernel functions are equal. Therefore for the choice of parameter  $c$  we will use a condition of a maximum of likelihood function [2]

$$L(c) = \prod_{j=1}^n \bar{p}(x^j),$$

where

$$\bar{p}(x^j) = \frac{1}{(n-1)c} \sum_{\substack{i=1 \\ i \neq j}}^n \Phi\left(\frac{x^j - x^i}{c}\right).$$

We will carry out the choice of value of parameter  $\beta$  according to expression  $\beta = \Delta/(2N)$ , where  $\Delta$  - a range of values of a random value  $x$ , and  $N$  - quantity of intervals of its sampling which are determined, for example, by Sturges' rule [8]

$$N = \log_2 n + 1 \quad (10)$$

or Heinhold-Gaede rule [9]

$$N = \sqrt{n}. \quad (11)$$

## II. ASYMPTOTIC PROPERTIES OF NONPARAMETRIC PROBABILITY DENSITY ESTIMATION

Let's give asymptotic properties of one-dimensional nonparametric assessment of a probability density (3) in case the required probability density  $p(x)$  decays in a row Taylor in each point  $x$ . When performing transformations we will use the technique offered in work [3].

For definition of conditions of an asymptotic unbiasedness of nonparametric assessment  $\bar{p}_1(x)$  we will calculate its expectation

$$\begin{aligned} M(\bar{p}_1(x)) &= \frac{1}{2\beta n c} \sum_{i=1}^n \int_{x-\beta}^{x+\beta} \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^i}{c}\right) du p(x^i) dx^i = \\ &= \frac{1}{2\beta c} \int_{x-\beta}^{x+\beta} \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-t}{c}\right) du p(t) dt, \end{aligned}$$

where observations  $x^i, i=1, \dots, n$  are characterized by the same probability density  $p(t)$ .

Let's carry out serial changes of variables (at first  $u = x + z\beta$ , and then  $(x + z\beta - t)/c = u_1$ ), we will receive

$$\begin{aligned} M(\bar{p}_1(x)) &= \frac{1}{2c} \int_{-1}^1 \int_{-1}^1 \Phi\left(\frac{x+z\beta-t}{c}\right) p(t) dt dz = \\ &= \frac{1}{2} \int_{-1}^1 \int_{-1}^1 \Phi(u_1) p(x+z\beta-cu_1) du_1 dz. \end{aligned} \quad (12)$$

Let's spread out function  $p(x+z\beta-cu_1)$  in a row of Taylor in a point  $x$  and we will integrate a right member of expression (12) taking into account properties of kernel function. As a result we will receive asymptotic expression

$$M(\bar{p}_1(x)) \sim p(x) + \frac{1}{2} \left( \frac{\beta^2}{3} + c^2 \right) p^{(2)}(x), \quad (13)$$

where  $p^{(2)}(x)$  - probability density flexon  $p(x)$ .

Let's calculate an average quadratic error of approximation

$$M(\bar{p}_1(x) - p(x))^2 = M(\bar{p}_1^2(x)) - 2p(x)M(\bar{p}_1(x)) + p^2(x). \quad (14)$$

Let's define the first item in a right member (14)

$$\begin{aligned} M(\bar{p}_1^2(x)) &= \frac{1}{4n^2 c^2 \beta^2} \left[ \sum_{i=1}^n \left( \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^i}{c}\right) du \right)^2 p(x^i) dx^i + \right. \\ &+ \left. \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^i}{c}\right) du \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-x^j}{c}\right) du p(x^i) p(x^j) dx^i dx^j \right] = \\ &= \frac{1}{4n^2 c^2 \beta^2} \left[ n \int_{x-\beta}^{x+\beta} \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-t}{c}\right) du p(t) dt + \right. \\ &+ \left. n(n-1) \left( \int_{x-\beta}^{x+\beta} \int_{x-\beta}^{x+\beta} \Phi\left(\frac{u-t}{c}\right) du p(t) dt \right)^2 \right]. \end{aligned}$$

After transformations on above to the given technique, we have

$$M(\bar{p}_1^2(x)) \sim p^2(x) + \frac{p(x)}{nc} \int \Phi^2(u) du + \left( \frac{\beta^2}{3} + c^2 \right) p^{(2)}(x) p(x) + \frac{1}{4} (p^{(2)}(x))^2 \left( \frac{\beta^2}{3} + c^2 \right)^2. \quad (15)$$

When obtaining this result the composed smallness  $0(c/n)$ ,  $0(\beta^2/(nc))$  and more were not considered.

Let's substitute in (14) expression (13) and (15), we will receive [7]

$$M(\bar{p}_1(x) - p(x))^2 = \frac{p(x)}{nc} \int \Phi^2(u) du + \frac{1}{4} (p^{(2)}(x))^2 \left( \frac{\beta^2}{3} + c^2 \right)^2. \quad (15)$$

From the analysis of expressions (12), (15) follows that when performing conditions  $\beta(n) \rightarrow 0$ ,  $c(n) \rightarrow 0$ ,  $nc(n) \rightarrow \infty$  for  $n \rightarrow \infty$  nonparametric probability density estimation  $\bar{p}_1(x)$  has properties of an asymptotic unbiasedness and convergence in mean squared and consequently, is a consistent estimate  $p(x)$ .

### III. ANALYSIS OF COMPUTING EXPERIMENTS RESULTS

Let's compare approximating properties of traditional nonparametric assessment of a probability density (1) and its modification (3) at restitution  $p(x)$  to the uniform and normal distribution laws of a random value.

At synthesis of statistics  $\bar{p}(x)$  function (5) and an optimum kernel of V. A. Epanechnikov were used (7). Formation of nonparametric probability density estimation (3) was carried out on the basis of kernel functions (6) and (8). Optimization  $\bar{p}_1(x)$  was performed according to the technique given above. As an efficiency factor of the compared statisticians estimates of their shift  $\bar{W}_1$ , a mean squared deviation  $\bar{W}_2$  and dispersion  $\bar{W}_3$  were used. Computing experiments when determining values  $\bar{W}_j$ ,  $j = \overline{1, 3}$  in specific conditions were conducted 50 times with the subsequent their averaging. At the same time the volume of control selection was accepted equal 1000 which elements were formed according to a probability density  $p(x)$ .

Let's consider results of a research at restitution of a probability density of a random value with the uniform distribution law.

With increase in number  $n$  input statistical datas of value  $\bar{W}_j$ ,  $j = \overline{1, 3}$  decrease for the studied statisticians that will be

coordinated with results of the analysis of their asymptotic properties. The offered probability density estimation  $\bar{p}_1(x)$  in comparison  $\bar{p}(x)$  has by (1.5 – 2) times smaller dispersion and mean squared deviations in conditions when at their synthesis kernel (6) and (8) functions, and also a formula of sampling are used (10). At the same time the shift of the  $\bar{W}_1(\bar{p}_1(x))$  statistics  $\bar{p}_1(x)$  from  $p(x)$  significantly more values  $\bar{W}_1(\bar{p}(x))$  for probability density assessment  $\bar{p}(x)$ .

If at construction  $\bar{p}(x)$  the kernel function of V. A. Epanechnikov is applied (7), that is observed decrease by (10 – 15) % of values of estimates of its indexes  $\bar{W}_j$ ,  $j = \overline{1, 3}$ . However, the regularities noted above generally remain.

Application of a formula of sampling of Heinhald-Gaede (11) allows to reduce shift  $\bar{W}_1(\bar{p}_1(x), 11)$  nonparametric assessment of a probability density  $\bar{p}_1(x)$  to 30% in comparison with results  $\bar{W}_1(\bar{p}_1(x), 10)$  of use of the Sturges' rule (10).

At the same time dispersion  $\bar{W}_3(\bar{p}_1(x), 11)$  statistics  $\bar{p}_1(x)$  increases by 20% of value  $\bar{W}_3(\bar{p}_1(x), 10)$  that is especially shown at great values  $n$ . However in these conditions dispersion and a mean squared deviation of nonparametric assessment  $\bar{p}_1(x)$  are less than for traditional statistics  $\bar{p}(x)$ .

At estimation of the normal distribution law approximating properties of statisticians  $\bar{p}(x)$  and  $\bar{p}_1(x)$  significantly decrease. This fact is explained by nonuniformity of value distribution of a random value that leads to decrease of best values of coefficients of a diffuseness and increase in dispersion of statisticians  $\bar{p}(x)$ ,  $\bar{p}_1(x)$ . Noted regularity especially is shown when using the procedure of sampling Heinhald and Gaede. However advantage of nonparametric assessment of a probability density  $\bar{p}_1(x)$  over  $\bar{p}(x)$  remains when using nuclear functions (5), (7) and formulas of sampling (10), (11).

### ACKNOWLEDGMENT

Use of the smoothing operator at synthesis of nonparametric probability density estimation like Rosenblatt-Parzen allows receiving its modifications with the increased approximating properties. The offered modifications unlike traditional nonparametric assessment of a probability density differ implicitly in the given form of kernel function. This feature allows changing not only definition range of kernel function, but also its look by their optimization. The procedure of optimization of the studied statisticians is implemented by the serial choice of coefficient of a diffuseness of kernel function from a condition of a maximum of function of credibility and parameter of the smoothing operator. At the same time formulas of sampling of an interval of values of a

random value are used. In comparison with nonparametric assessment of a probability density of Rosenblatt-Parzen the studied its modification is characterized by great value of shift, but smaller dispersion and a mean squared deviation. Noted properties remain under much different conditions of computing experiments: a type of the restored probability density and kernel function, and also the used rules of sampling of an interval of values of a random value.

The received results create a basis of creation of nonparametric estimates of many-dimensional probability densities from implicitly given kernel function, and also for development of analytical criteria of quality of approximation. The studied estimates of a probability density are important at creation of nonparametric algorithms of information processing and a decision making in the conditions of statistical data of restricted volume.

#### REFERENCES

- [1] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statistic.*, vol. 33, pp. 1065-1076, 1962.
- [2] L. Devroyes and L. Györfi, *Nonparametric Density Estimation: The L1 View*. Wiley, New York, 1984.
- [3] V.A. Epanechnikov, "Nonparametric estimator of multidimensional probability density," *Teor. Veroyatn. Ee Primen.*, vol. 14, Iss. 1, pp. 156–161, 1969.
- [4] A.V. Lapko and V.A. Lapko, "Regression estimate of the multidimensional probability density and its properties," *Opt. Instrum. Data Proc.*, vol. 50, no. 2, pp.148-153, 2014.
- [5] A.V. Lapko and V.A. Lapko, "Comparison of the effectiveness of methods for sampling the range of variation of random quantities in synthesis of nonparametric estimates of probability density," *Measur. Techn.*, vol. 57, no. 3, pp. 222 – 227, 2014.
- [6] A.V. Lapko and V.A. Lapko, "Construction of confidence limits for the probability density function on the basis of nonparametric estimation of the function," *Measur. Techn.*, vol. 56, no. 12, pp. 1354 – 1357, 2014.
- [7] A.V. Lapko and V.A. Lapko, "Nonparametric estimate of a parzen-type probability density with an implicitly specified form of the kernel," *Measur. Techn.*, vol. 59, no. 6, pp. 571 – 576, 2016.
- [8] H.A. Sturges, "The Choice of a Class Interval," *J. Amer. Stat. Association*, vol. 21, no. 153, pp.65–66, 1926.
- [9] I. Heinhold and K. W. Gaede, *Ingenieur Statistik*, Springer-Verlag, Munich, Vienna, 1964.