

УДК 519.24

Piecewise Polynomial Models for Aggregation and Regression Analysis in Remote Sensing of the Earth Problems

Olga A. Popova*

*Siberian Federal University
79 Svobodny, Krasnoyarsk, 660041, Russia*

Received 27.12.2017, received in revised form 19.08.2018, accepted 21.10.2018

We discuss the procedures of data aggregation as a preprocessing stage for subsequent to regression modeling. An important feature of study is demonstration of the way how represent the aggregated data. It is proposed to use piecewise polynomial models, including spline aggregate functions. We show that the proposed approach to data aggregation can be interpreted as the frequency distribution. To study its properties density function concept is used. Applying data aggregation models as input and output variables we propose a new probability density function value linear regression model (Distributions Regression). To calculate the data aggregation and regression model we employ numerical probabilistic analysis (NPA). To demonstrate the degree of the correspondence of the proposed methods to reality, we developed a theoretical framework and considered numerical examples.

Keywords: numerical probabilistic analysis, data aggregation, regression modeling, piecewise polynomial model, density function.

Citation: Popova O.A. Piecewise polynomial models for aggregation and regression analysis in remote sensing of the Earth problems, J. Sib. Fed. Univ. Eng. technol., 2018, 11(8), 964-973. DOI: 10.17516/1999-494X-0118.

Кусочно-полиномиальные модели для агрегации и регрессионного анализа в задачах дистанционного зондирования Земли

О.А. Попова

*Сибирский федеральный университет
Россия, 660041, Красноярск, пр. Свободный, 79*

Предложены новые подходы для исследования и анализа зависимостей в эмпирических данных. Обсуждаются вопросы агрегирования и различные виды математических моделей

© Siberian Federal University. All rights reserved

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

* Corresponding author E-mail address: BDoBronets@yandex.ru

агрегированных данных. Для больших объемов данных предлагается использовать процедуры агрегирования на основе кусочно-полиномиальных моделей. Рассматриваются вопросы повышения точности построения кусочно-полиномиальных моделей в виде полиномиальных сплайнов. Рассмотрены новые подходы в задачах восстановления функциональных зависимостей на основе сплайн-агрегаций.

Ключевые слова: численный вероятностный анализ, агрегация, регрессионный анализ, кусочно-полиномиальные модели, функции плотности.

1. Introduction

The linear regression analysis is often used to discovery dependencies of the empirical data. Properties of empirical data including the type and level of uncertainty significantly affect on the results of the simulation. It is well-known, that the random uncertainty concept addressed by probability theory plays a fundamental role to study data uncertainty.

This work continues the research begun in [1]. As a new mathematical branch, the linear regression for interval-valued and histogram-valued variables presents a new study for uncertainty data processing [2-5].

This problem becomes more complicated if large amounts of data are processed. In this case it is useful to look at the empirical data in an aggregated form. Aggregation is a popular method of converting data. For example, the application of the histogram allows to reduce dimension of data set and level of uncertainty and to significantly increase the efficiency of numerical calculations. It is important to note that histograms are the examples of symbolic data used in the Symbolic data analysis [3]. Symbolic Data Analysis and Data Mining use histograms to study a variety of different processes and are applied for modeling the variability of quantitative characteristics. Histogram data models and histogram regression models based on the Symbolic analysis is a new important direction to discover knowledge in a data base.

In our study we consider a new approach to regression modeling using input data aggregation. To develop our approach for performing efficient aggregation we employ piecewise polynomial aggregation functions, including piecewise linear functions and piecewise constant functions. Histogram is a good example of piecewise constant functions of which are perfectly employed in our study.

To examine the structure of data aggregation we use the probability density functions (PDF). The concept of the mathematical aggregation functions is used to the regression modeling. To illustrate this we will regard the spline aggregation function in more detail. This approach will allow to employ the density function models as input and output data. It is of further importance that, the data uncertainty is studied to identify the relationship between the input and output characteristics when the input probability density functions are unknown. Thus, in order to describe any specific PDF we need to consider their spline interpolation.

In this work we propose a new linear regression model named a PDF-valued variable regression. The abbreviated form of the regression model is called a Distributions Regression. If we use a spline aggregation model it is named a PDF-spline valued variable Regression Model and a Distributions Regression in shot. The following statements confirm the justification of PDF-spline models. The application of the spline procedures allows big data aggregating reduce the level of uncertainty and

to significantly increase the efficiency of numerical calculations. These splines allow considerably accurate representations of the arbitrary distribution.

To demonstrate the degree of the relevance of the proposed methods to reality, we developed a theoretical study and provided numerical examples to illustrate it. With this we propose a conclusive discussion of this approach applicability to the uncertainty treatment and big data processing. The comparison of NPA and Monte Carlo method showed good agreement of the results. At the same time, numerical experiments demonstrate that the PDF arithmetic is more than hundred times faster than the Monte Carlo method [6]. As a result, the NPA approach can be successfully applied for solving computational and engineering problems.

The structure of the remaining sections is stated as follows. Section 2 reviews the data aggregation models. Section 3 covers the discussion the numerical arithmetic for probability density function. In Section 4, we review the spline interpolation. In Section 5 we study the Spline aggregation. The questions about the application of the regression approach to spline-aggregated time series are discussed in Section 6, Section 7 concludes the paper.

2. Data aggregation

In this section we will look at data aggregation as a pretreatment method for subsequent for numerical modeling.

The essence of the aggregation procedure are methods for reducing the dimension of the original empirical data, knowledge discovery and reduce data uncertainty. Data aggregation procedure plays the most concerned role in the process of to extracting useful information from a large volume of data. The essence of the aggregation procedure is to constitute methods for reducing the initial data set to a less data collection. Aggregation can be considered as a data conversion process with a high degree of detail to a more generalized representation. An example of such procedures is a simple summation, calculation of the average, median, mode and range of maximum or minimum values.

The aggregation procedure has its own advantages and disadvantages. On the positive side, we note, that the detailed data are often very volatile due to the impact of different random factors, making it difficult to discover general trends and data patterns. In many cases it is useful first to look at numerical big data in an aggregated form such as a summation or an average.

It is important to bear in mind that the use of such aggregation procedures as averaging, exclusion of the extreme values (emission) and, smoothing procedure can lead to a loss of important information. There are various methods of data aggregation. Therefore, the choice of the method of aggregation is a complex problem, as wrongly selected numerical methods of calculation may introduce additional uncertainty that was not presented in the original problem.

The data aggregation can use various mathematical models. NPA outlines the following key models histograms, frequency polygons and splines.

Histograms. A random variable whose probability density function is represented by a piecewise constant function called a histogram. Any histogram P is defined by a grid $\{x_i \mid i = 0, \dots, n\}$ and a set of values p_i , $i = 1, \dots, n$, such that the histogram takes the constant value p_i at the interval $[x_{i-1}, x_i]$.

Histograms are widely used for the processing and analysis of remote sensing data. For example, in [?] is considered the problem of the study of natural processes on the basis of space and ground

monitoring data. In addition to histograms we discuss the piecewise linear functions (polygons) and splines.

The piecewise linear function (Frequency Polygons). Piecewise linear functions can be considered as a tool of approximation of the density function of a random variable. A piecewise linear function is a function composed of straight-line sections. The frequency polygon (FP) is a continuous density estimator based on the histogram. In one dimension, the frequency polygon is the linear interpolant of the midpoints of an equally spaced histogram.

These are the simplest splines. Although such functions are relatively simple, they have good approximating properties, and they can be tools for approximating probability density functions.

Spline. A spline is a sufficiently smooth polynomial function that is piecewise-defined, and possesses a high degree of smoothness at the places where the polynomial pieces connect (which are known as knots). We will consider the probability density of the random variables as an approximated spline.

3. Numerical arithmetic for Probability Density Function

To adapt the classical regression model to these kinds of data aggregation considered as PDF-valued variables we must use relevant numerical arithmetic. To meet this aim, we will consider in more detail the numerical operations on probability density functions developed in the framework of numerical probability analysis.

Let a system of two continuous random variables (x_1, x_2) with probability density function be $p(x_1, x_2)$. The densities resulting from arithmetic operations on random variables following those distributions are given by [6]. For example, to find the probability density function $p_{x_1+x_2}$ of a sum of two random variables $x_1 + x_2$ we use by

$$p_{x+y}(x) = \int_{-\infty}^{\infty} p(x-v, v)dv = \int_{-\infty}^{\infty} p(v, x-v)dv. \quad (1)$$

To compute of the probability density function $p_{x/y}$ as the result of dividing of two random variables x_1/x_2 we use

$$p_{x/y} = \int_0^{\infty} vp(xv, v)dv - \int_{-\infty}^0 vp(v, xv)dv. \quad (2)$$

The probability density function p_{xy} resulting from multiplying of two random variables xy is computed accordingly by

$$p_{xy} = \int_0^{\infty} (1/v)p(x/v, v)dv - \int_{-\infty}^0 (1/v)p(v, x/v)dv. \quad (3)$$

Commutativity and associativity of arithmetic operations on addition and multiplication directly follows from these expressions. Let us consider the question of the existence of inverse elements in summa and multiplication operations. Note that there is a solution to the equation $a + x = 0$ which could be represented as a joint probability density function of (a, x)

$$p_{-a}(x, y) = \begin{cases} p_a(x) & \text{if } x = -y; \\ 0 & \text{if } x \neq -y, \end{cases}$$

where $p_a(x)$ is probability density function for random variable a .

There by $p_{-a}(x, y)$ show the inverse elements in summa operations for a . Similarly, we can easily construct the inverse element for multiplication. Let $f(x)$ be probability density function for a random variable x . Then

$$f_a(x) = \frac{1}{a} f\left(\frac{x}{a}\right) \tag{4}$$

is probability density function for random variable ax , where a is real variable.

We consider the problem of calculating by the derivative with respect to the parameter a for ax . It follows directly from the definition of the derivative that

$$\frac{d}{da} ax = \lim_{h \rightarrow 0} \frac{(a+h)x - ax}{h} = \frac{(a+h-a)x}{h} = x.$$

Consider the case of computing the derivative with respect to the parameter a from $f_a(x)$

$$\frac{d}{da} f_a(x) = \frac{d}{da} \left(\frac{1}{a} f\left(\frac{x}{a}\right) \right) = -\frac{1}{a^2} f\left(\frac{x}{a}\right) - \frac{1}{a} f'\left(\frac{x}{a}\right) \frac{x}{a^2}. \tag{5}$$

Next, we calculate the derivative with respect to a for probability density function f_{ax+y} of summa $ax + y$.

$$\begin{aligned} \frac{d}{da} f_{ax+y}(z) &= \frac{d}{da} \int_{-\infty}^{\infty} \left(\frac{1}{a} f_1\left(\frac{\xi}{a}\right) \right) f_2(z - \xi) d\xi = \\ &= \int_{-\infty}^{\infty} \left(-\frac{1}{a^2} f_1\left(\frac{\xi}{a}\right) - \frac{1}{a} f_1'\left(\frac{\xi}{a}\right) \frac{\xi}{a^2} \right) f_2(z - \xi) d\xi. \end{aligned} \tag{6}$$

Let f_1 be probability density function of x , where f_2 is probability density function of y . If random variables (x, y) are independent, we can calculate the joint probability density function $p(x, y)$ as a product of $p(x, y) = p_1(x)p_2(y)$, where p_1, p_2 will be a probability density function of x, y .

Let p_i be represented to piecewise polynomial functions belonging to the mesh $\omega_i = \{z_l^{(i)}, l = 0, \dots, n^{(i)}\}$ and function p_i is the polynomial $s_k^{(i)}$ on the each interval of $[z_k^{(i)}, z_{k+1}^{(i)}]$.

Consequently $p(x, y)$ is a polynomial $s_k^{(1)}(x_1)s_l^{(2)}(x_2)$ on $[z_k^{(1)}, z_{k+1}^{(1)}] \times [z_l^{(2)}, z_{l+1}^{(2)}]$.

For example, we consider to build the density probability function p_{x+y} for sum of two variables $x + y$.

Let it be necessary to calculate the value of the probability density function $p_{x+y}(z)$ from the sum of two variables at a certain point x . The integral (1) will be calculated numerically. Then we define the indexes $(k, l) \in I_x$ of cells $[z_k^{(1)}, z_{k+1}^{(1)}] \times [z_l^{(2)}, z_{l+1}^{(2)}]$ that intersect the line $x + y = z$. On each such cell, the calculation of the integral (1) reduces to calculating the integral

$$Int_{(k,l)} = \int_{z_k^{(*)}}^{z_{k+1}^{(*)}} s_k^{(1)}(v)s_l^{(2)}(x-v)dv,$$

where $z_k^{(*)}, z_{k+1}^{(*)}$ are the coordinates of the projection on the axis x_1 of the segment

$$[z_k^{(1)}, z_{k+1}^{(1)}] \times [z_l^{(2)}, z_{l+1}^{(2)}] \cap \{x_1 + x_2 = x\}.$$

Since the integrand is the product of polynomials, we can calculate $Int_{(k,l)}$ both analytically and by using the numerical integration procedures that are exact on polynomials.

Finally, we obtain the following form

$$p_{x+y}(z) = \sum_{(k,l) \in I_x} Int_{(k,l)}.$$

To construct a piecewise polynomial approximation of p_{x+y} , we construct a mesh $\omega = \{x_0, x_1, \dots, x_n\}$ in the support domain to compute $f_i = p_{x+y}(x_i)$. Next, using the values f_i on the mesh ω , we construct, for example, a cubic spline s . In this case the estimate is calculated as

$$\|p_{x+y}^{(v)} - s^{(v)}\| \leq Kh^{4-v} \|p_{x+y}^{(4)}\|, v = 1, 2, 3.$$

The computed value will then look as follows

$$\text{norm} = \int s(x) dx,$$

If the norm $\neq 1$, then we finally assume that $s(x) := s(x)/\text{norm}$.

4. Spline interpolation

Prior to consideration of spline aggregation, we propose study of mathematical models applicable for the representation of splines and will discuss the interpolation questions with their application.

Let $\omega = \{x_0 < x_1 < x_2 < \dots < x_n\}$ be mesh and interpolation conditions

$$s(x_i) = f(x_i), i = 0, \dots, n.$$

Where boundary conditions are given as

$$s'(x_0) = 0, s'(x_n) = 0.$$

The cubic spline on a mesh $\{x_i\}$ with step $\bar{h} = \max(x_{i+1} - x_i), i = 0, \dots, n$ satisfies the estimate

$$\|f^v - s^v\| \leq \bar{h}^{4-v} \|f^{(4)}\|, v = 0, 1, 2. \tag{7}$$

The task of the spline construction is reduced for solving a system of linear algebraic equations with a tridiagonal matrix [8]

$$\lambda_j m_{j-1} + 2m_j + \mu_j m_{j+1} = d_j, \tag{8}$$

$$2m_0 + m_1 = 3(f_1 - f_0)/h_1 - h_1 z_0^2/2,$$

$$2m_n + m_{n-1} = 3(f_n - f_{n-1})/h_n + h_n z_n^2/2,$$

$$d_j = 3\lambda_j(f_j - f_{j-1})/h_j + 3\mu_j(f_{j+1} - f_j)/h_{j+1}, \quad j = 1, \dots, N-1,$$

where $m_i = s'(x_i)$. As a result, a cubic spline on the intervals $[x_{j-1}, x_j], j = 1, \dots, N$ will have the following representation [8]:

$$s(x) = m_{j-1}(x_j - x)^2(x - x_{j-1})/h_j^2 - m_j(x - x_{j-1})^2(x_j - x)/h_j^2 + \tag{9}$$

$$+ f_{j-1}(x_j - x)^2(2(x - x_{j-1}) + h_j)/h_j^3 + f_j(x - x_{j-1})^2(2(x_j - x) + h_j)/h_j^3,$$

5. Spline aggregation

Let consider the spline approach to build regression model with the Distributions-valued variables. This approach is useful due to the following reasons. Underlying of this approach is the notion of the spline. The spline can be regarded as a mathematical object that is easy to describe and calculate the mathematical procedures and operations, in the process of maintaining the essence of data frequency distribution.

Since the spline is a piecewise polynomial function then it can be regarded as a data aggregation function in aggregation issues. Aggregation function performs numerical calculations on a data set and returns the spline values. Splines are useful for data uncertainty analysis due to fact that they adequately represent the random distribution of random variables.

Despite its simplicity, the spline also covers all possible ranges of probability density function estimation. Simple and flexible spline structure greatly simplifies their use in numerical calculations and it has a clear visual image, which is useful for analytical conclusions. It is important to note that the construction of regression models with aggregated inputs require the use of appropriate numerical procedures. To this end, we consider numerical probabilistic analysis. We propose to use of the numerical probabilistic analysis to compute the arithmetic operations for the aggregated data and to apply for regression modeling.

Let us assume that we know the samples $\Xi = \{\xi_1, \xi_2, \dots, \xi_N\}$ of a random variable ξ with probability density function $f(x)$ and support $[a, b]$. Далее рассмотрим использование Richardson's extrapolation для повышение точности kernel estimator.

The basic kernel estimator may be written compactly as [7]

$$\hat{f}^h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - \xi_i}{h}\right) = \frac{1}{Nh} \sum_{i=1}^N K_h(x - \xi_i),$$

where $K_h(t) = K(t/h)/h$.

Note

$$K_h(x, \xi_i) = K\left(\frac{x - \xi_i}{h}\right),$$

where ξ is a random variable with probability density function $f(x)$.

Then

$$E[\hat{f}^h(x)] = E[K_h(x, \xi)]$$

and

$$\sigma_N = \text{Var}[\hat{f}^h(x)] = \frac{1}{N} \text{Var}[K_h(x, \xi)].$$

The value of the mathematical expectation can be written as

$$E[K_h(x, \xi)] = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x-t}{h}\right) f(t) dt = \int_{-\infty}^{\infty} K(\eta) f(x - h\eta) d\eta.$$

Notice, that

$$f(x - h\eta) = f(x) - hf'(x)\eta + \frac{h^2}{2} f''(x)\eta^2 + \frac{h^3}{6} f^{(3)}(x)\eta^3 + O(h^4).$$

$$\begin{aligned} E[K_h(x, \xi)] &= f(x) \int_{-\infty}^{\infty} K(\eta) d\eta - hf'(x) \int_{-\infty}^{\infty} \eta K(\eta) d\eta + \\ &+ \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} K(\eta) \eta^2 d\eta + \frac{h^3}{6} f^{(3)}(x) \int_{-\infty}^{\infty} K(\eta) \eta^3 d\eta + O(h^4). \end{aligned}$$

Suppose that the kernel K satisfies the requirements

$$\int_{-\infty}^{\infty} K(\eta) d\eta = 1, \int_{-\infty}^{\infty} \eta K(\eta) d\eta = 0$$

and

$$\int_{-\infty}^{\infty} \eta^3 K(\eta) d\eta = 0.$$

Denote

$$\int_{-\infty}^{\infty} \eta^2 K(\eta) d\eta = \sigma^2.$$

Then

$$\bar{f} = E[K_h(x, \xi)] = f(x) + \sigma^2 h^2 f''(x)/2 + O(h^4)$$

and

$$E[\hat{f}^h(x) - f(x)] = \sigma^2 h^2 f''(x)/2 + O(h^4).$$

Define $f^h(x)$

$$f^h(x) = E[\hat{f}^h(x)] = f(x) + \sigma^2 h^2 f''(x)/2 + O(h^4). \quad (10)$$

and $f^{2h}(x)$

$$f^{2h}(x) = E[\hat{f}^{2h}(x)] = f(x) + 4\sigma^2 h^2 f''(x)/2 + O(h^4). \quad (11)$$

Let us apply the Richardson's extrapolation to $f^h(x)$ and $f^{2h}(x)$ [9]. In the next stage, we multiply (10) on 1/4 to subtract the result from (11) Excluding $\sigma^2 h^2 f''(x)/2$ from (10) and (11), we get

$$f(x) = \frac{4}{3} f^h(x) - \frac{1}{3} f^{2h}(x) + O(h^4).$$

Noting that we have constructed the approximation to the function $f(x)$

$$f_{cor}^h(x) = \frac{4}{3} \hat{f}^h(x) - \frac{1}{3} \hat{f}^{2h}(x), \quad (12)$$

with the accuracy $O(h^4)$.

Thus, successively assuming $z \in \omega$, we obtain the values $f_{cor}^h(x_i) = f(x_i) + O(h^4)$ and the system of linear algebraic equations (8) for constructing a cubic spline. To improve the reconstruction accuracy of probability density at the point z we use the combination of kernel assessments with the parameters h and $2h$.

As an evidence we refer to Schweizer who states, that "distributions are the numbers of the future" [10]. Thus, instead of simplifying them, it seems better to propose methods which deal with distributions directly. In order to do this, one has to determine how to represent the observed distributions.

In our study we propose to represent them by using a piecewise polynomial aggregation function, as long as it offers a good tradeoff between simplicity and accuracy.

6. Distributions Regression

Consider a linear model

$$Y = a_0 + \sum_{i=1}^n a_i X_i + \varepsilon,$$

where X_i , $i = 1, \dots, n$ are independent predictor variables, Y is a dependent variable, ε is an error. From the observed values of Y_j $X_{i,j}$ after the aggregation of the density Y , X_i are represented by splines: S_y , S_i .

We shall seek the unknown parameters a_i , $i = 0, 1, \dots, n$ starting from the minimum of the functional

$$\Phi(a_0, a_1, \dots, a_n) = \|S_y - (a_0 + \sum_{i=1}^n a_i X_i)\|_2 \rightarrow \min.$$

By virtue of the independence of X_i , numerical operations on density functions can be used to calculate the functional $\Phi(a_0, a_1, \dots, a_n)$. The minimization of the functional $\Phi(a_0, a_1, \dots, a_n)$ can be carried out by the method of steepest descent.

Numerical example. Let us consider model problem

$$Y = a_0 + \sum_{i=1}^n a_i X_i + \varepsilon,$$

For numerical realization, X_1 , X_2 were generated as sums of random variables with an Irvine-Hall distribution $n = 3$ and shifted by 1 and 2, respectively, ε with probability density function $(|2x| - 1)^2 (2|2x| + 1)$ with support $[-0.5, 0.5]$.

The variable Y was constructed as follows $Y = X_1 + X_2 + \varepsilon$.

The minimization of the functional $\Phi(a_0, a_1, a_2)$ was carried out by the method of steepest descent. For $a_0 = -0.089$, $a_1 = 1.031$, $a_2 = 1.029$, the value $\Phi(a_0, a_1, a_2)$ did not exceed the value $0.3 \cdot 10^{-3}$.

Thus, a numerical example showed the possibility of using distributions regression.

7. Conclusion

Data transformation during the data aggregation phase is an important direction in the analysis of data. Well-chosen data models at the aggregation stage allow determine the form of the input variables and select the appropriate procedures and arithmetic for later modeling.

The use of regression modeling on the basis of piecewise polynomial models opens up new possibilities in forecasting the problems of hydrology, remote sensing of the Earth, estimating the reliability of critical equipment.

References

- [1] Dobronets B.S., Popova O.A. The numerical probabilistic approach to the processing and presentation of remote monitoring data. *Journal of Siberian Federal University. Engineering & Technologies*, 2016, 9(7), 960-971.
- [2] Добронетц Б.С., Попова О.А. *Численный вероятностный анализ неопределенных данных*. Красноярск, Сибирский федеральный университет, Институт космических и информационных технологий. 2014. 168 с. [Dobronets B.S., Popova O.A. *Numerical probabilistic analysis of uncertain data*. Krasnoyarsk, Siberian Federal University, Institute of Space and Information Technologies. 2014. 168 p. (in Russian)]
- [3] Billard L., Diday E. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Chichester, John Wiley & Sons, Ltd., 2006. 321 p.
- [4] Dias S., Brito P. Linear Regression Model with Histogram-Valued Variables. *Published online in Wiley Online Library (wileyonlinelibrary.com)*. 2015, DOI:10.1002/sam.11260
- [5] Koenker R. *Quantile regression*. Cambridge university press, 2005. 349 p.
- [6] Dobronets B.S., Krantsevich A.M., Krantsevich N.M. Software implementation of numerical operations on random variables. *Journal of Siberian Federal University. Mathematics & Physics*, 2013, 6(2), 168–173.
- [7] Scott R.W. *Multivariate density estimation: theory, practice, and visualization*. New York, John Wiley & Sons, 2015. 380 p.
- [8] Ahlberg J.H., Nilson E.N., Walsh J.L. *The theory of splines and their applications*. New York, Academic Press, 1967. 284 p.
- [9] Dobronets B.S., Popova O.A. Improving the accuracy of the probability density function estimation. *Journal of Siberian Federal University. Mathematics and Physics*, 2017, 10(1), 16-21.
- [10] Schweizer B. Distributions Are the Numbers of the Future. *Proceedings of The Mathematics of Fuzzy Systems Meeting*, eds. A. di Nola and A. Ventre, Naples, University of Naples, 1984, 137–149.