Розглянуто задачу регресійного аналізу з нечітко заданими змінними. Сформульовано та обґрунтовано критерій якості оцінки регресійних коефіцієнтів, що враховує суттєві відмінності у точності завдання змінних. Запропоновано метод розв'язання задачі. Розглянуто і вирішено задачу нечіткої компараторної ідентифікації, коли значення змінної, яка пояснюється, не визначено, але можуть бути ранжовані за зменшенням будь-якого обраного показника

Ключові слова: нечіткий регресійний аналіз, нечіткі вихідні дані, нечітка компараторна ідентифікація

Рассмотрена задача регрессионного анализа с нечётко заданными переменными. Сформулирован и обоснован критерий качества оценки регрессионных коэффициентов, учитывающий существенные различия в точности задания переменных. Предложен метод решения задачи. Рассмотрена и решена задача нечёткой компараторной идентификации, когда значения объясняемой переменной не определены, но могут быть ранжированы по убыванию какого-либо выбранного показателя

Ключевые слова: нечёткий регрессионный анализ, нечёткие исходные данные, нечёткая компараторная идентификация

UDC 519.24
DOI: 10.15587/1729-4061.2017.107536

MODELS AND METHODS OF REGRESSION ANALYSIS UNDER CONDITIONS OF FUZZY INITIAL DATA

L. Raskin
Doctor of Technical Sciences, Professor, Head of Department*
O. Sira
Doctor of Technical Sciences, Professor*
E-mail: topology@ukr.net
Y. Ivanchykhin
PhD, Associate Professor*
*Department of distributed information systems and cloud technologies
National Technical University "Kharkiv Polytechnic Institute"
Kyrpychova str., 2, Kharkiv, Ukraine, 61002

1. Introduction

Regression analysis is a powerful and effective statistical method of constructing mathematical models that describe the relationship between the indicator of the functioning of the analyzed system y and the conditioning, explanatory independent variables (factors) $F_1, F_2, ..., F_m$.

In order to reveal this connection, a series of experiments is conducted in which each experiment $\left(F_{j1}, F_{j2}, ..., F_{jm}\right)$ determines its corresponding result, i.e., the value of the dependent variable y_j, where $j = 1, 2, ..., n.$ The sought connection is usually described by the Kolmogorov-Gabor polynomial, which in the simplest case has the form

$$y_j = x_0 + F_{j1}x_1 + F_{j2}x_2 + ... + F_{jm}x_m + \varepsilon_j.$$

Here F_{ji} is the value of the i-th independent variable in the j-th experiment; $i = 0, 1, 2, ..., m,$ and $j = 1, 2, ..., n.$

In the matrix, the above relation has the form $FX = Y$, where

$$F = \begin{pmatrix} 1 & F_{11} & F_{12} & ... & F_{1m} \\ 1 & F_{21} & F_{22} & ... & F_{2m} \\ ... & ... & ... & ... & ... \\ 1 & F_{n1} & F_{n2} & ... & F_{nm} \end{pmatrix}, \quad X = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ ... \\ x_m \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix}.$$

In the canonical regression analysis, the following basic assumptions are made.

1. The values of the independent variables F_i are measured without errors, and $i = 1, 2, ..., m.$

2. The dependent variable y in each experiment is estimated with a random error ε_j, which is normally distributed with a zero mathematical expectation and a known variance σ^2.

3. The random errors ε_j in different experiments are not correlated.

In these proposals, estimates of the unknown coefficients $x_0, x_1, ..., x_m$ of the regression polynomial are obtained by the least square method (LSM), minimizing the sum of the squared deviations of the values of the resulting variable y_i from the corresponding values

$$\sum_{i=0}^{m} x_i F_{ji},$$

predicted by the model.

The solution of many practical problems becomes more complicated when the initial assumptions of the classical regression analysis are not true. Of particular interest are the problems in which the initial data are not clearly estimated.

2. Literature review and problem statement

Apparently, one of the first works in which the problem of regression analysis was formulated and solved, taking into account the vagueness of the initial data, was [1]. In this paper, a linear model of fuzzy regression with a clear set of values of variables and indistinctly defined regression coefficients with triangular symmetric membership functions is introduced. The task of finding the set of regression coefficients is reduced to the problem of linear programming. Further, works [1–9] develop the approach proposed in the original article [1]. The general scheme for solving the problem in the terms that are introduced in [2] has the following form. The linear model that connects the explanatory variables $F=\{F_1,F_2,...,F_k\}$ and the explained y has the form

$$y=\tilde{b}_0+\tilde{b}_1F_1+\tilde{b}_2F_2+...+\tilde{b}_kF_k,$$

where $\{F_i\},\quad i=1,2,...,k$ is a set of explanatory variables; y is the variable to be explained; $\{\tilde{b}_i\},\quad i=0,1,...,k$ is a set of fuzzy numbers with membership functions:

$$M(\tilde{b}_i)=\begin{cases}0, & \tilde{b}_i<m_i-c_i,\\ \dfrac{\tilde{b}_i-(m_i-c_i)}{c_i}, & m_i-c_i\le\tilde{b}_i<m_i,\\ \dfrac{(m_i+c_i)-\tilde{b}_i}{c_i}, & m_i\le\tilde{b}_i\le m_i+c_i,\\ 0, & \tilde{b}_i>m_i+c_i.\end{cases}$$

There are n experiments resulting in the sets $F=\{F_{ji}\}$, $(y_1,y_2,...,y_n)$.

The task is to find the parameters $(m_i,c_i),\quad i=0,1,...,k$ that minimize the objective function

$$L=nc_0+\sum_{i=0}^{k}\sum_{j=1}^{n}c_i\left|F_{ji}\right|$$

as well as satisfy the constraints

$$m_0-c_0+(m_1-c_1)F_{j1}+...+(m_k-c_k)F_{jk}\ge\min_j y_j$$

and

$$(m_0+c_0)+(m_1+c_1)F_{j1}+...+(m_k-c_k)F_{jk}\le\max_j y,$$

$$j=1,2,...,n.$$

The usual linear programming problem is obtained.

Further, studies [3, 4] introduce asymmetric functions of the fuzzy numbers $b_i,\quad i=1,2,...,n,$ and studies [5–7] introduce trapezoidal membership functions. Then, in [8, 9], a polynomial regression model is introduced, with the group-based method of arguments being used to estimate the coefficients.

Starting with [10], another approach to estimating regression coefficients, based on the use of the method of least squares, has been developed. Another improvement in the regression model is the assumption that all observed variables are fuzzy numbers [11–19]. In this case, different hypotheses are used about the form of the membership functions of fuzzy regression coefficients (triangular, trapezoidal, and Gaussian). However, in all cases the general scheme for solving the problem is the same.

A predictable membership function for the fuzzy value of the explained variable is formed, taking into account the nature of the introduced regression model and the functions of fuzzy coefficients belonging to it as well as using the Zade generalization principle and the rules for performing operations on fuzzy numbers. On the other hand, this same membership function is determined from the experimental data. These two membership functions are compared with each other, and the resulting "distance" between them is used in the least-squares procedure to find the parameters that determine the membership functions of the regression coefficients. For example, in [16], this procedure is realized as follows. A model membership function is introduced for the fuzzy value of the explained variable y, which has a Gaussian character

$$\mu(y)=\exp\left\{-\frac{(y-b)^2}{2c^2}\right\}.$$

The coefficients b and c are described by the regression models:

$$b=\sum_{i=1}^{m}b_iF_i,\ \text{and}\ c=\sum_{i=1}^{m}c_iF_i.$$

The "distance" between the model membership function $\mu_0(y)$ and the real membership function $\mu_p(y)$ is calculated by the formula

$$R=\sqrt{\frac{\int_{\underline{y}}^{\overline{y}}\left[\mu_0(y)-\mu_p(y)\right]^2}{\overline{y}-\underline{y}}dy},$$

where $\left[\underline{y},\overline{y}\right]$ is the range of the observed values of y.

Finally, in [20–22], the determination of the parameters of the fuzzy coefficients of the regression model is performed in the Chebyshev metric.

It should be noted that, despite a very large number of works on the problem of regression analysis (in particular, [22] contains a detailed and qualified review of the most significant and interesting results obtained during the period of 1982–2017), some important questions remain insufficiently researched. For example, the often used criterion for assessing the quality of a solution to the regression analysis problem cannot be considered convincingly justified. This criterion minimizes the differences between the model function of the fuzzy predicted value belonging to the explained variable and the membership function of this variable that is obtained after processing the experimental data. The fact is that the accuracy of estimating the values of independent and dependent variables can differ substantially, and it is for the worse with respect to the dependent variable. In this connection, the result of solving the problem does not necessarily provide a minimum of the mean total fuzziness of the predicted value of the explained variable. Apparently, a preferable solution would be the one that satisfies the following two natural requirements:

– proximity to the modal value of the fuzzy explained variable that is obtained during the processing of the experimental data;

– maximum compactness of the attribution function of the predicted value of the explained variable, taking into account the regression relationship.

A completely different type is the problem of estimating the coefficients of the regression model in the absence of data on the values of the dependent variable. Direct use of traditional regression analysis technologies is impossible if in the experiments all available information is limited only by data on the values of the independent variables (F_{ji}). A similar situation arises when comparing a set of similar objects by data on a set of their characteristics. Of course, this information is not sufficient for constructing the regression model

$$R(x)=x_0+x_1F_1+x_2F_2+...+x_mF_m, \quad (1)$$

where $F=(F_1,F_2,...,F_m)$ is a set of factors that are the characteristics of the object; $X=(x_0,x_1,...,x_m)$ is a set of weight coefficients.

However, the use of data on the characteristics of objects allows, for example, expert way, ranking the compared objects in a descending order of the value of their any resulting characteristics, for example, "usefulness". In this case, we obtain

$$R_1(X)>R_2(X)>...>R_{n-1}(X)>R_n(X). \quad (2)$$

This relation is used in the standard problem of comparator identification [23] for finding the coefficients of model (1). Inequalities (2) define the following system of strict inequalities:

$$\begin{aligned}&\zeta_1(X)=R_2(X)-R_1(X)=\\&=x_1(F_{21}-F_{11})+x_2(F_{22}-F_{12})+...+x_m(F_{2m}-F_{1m})=\sum_{i=1}^{m}x_iV_{1i}<0,\\&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\&\zeta_{n-1}(X)=R_n(X)-R_{n-1}(X)=\\&=x_1(F_{1n}-F_{1,n-1})+...+x_m(F_{mn}-F_{m,n-1})=\sum_{i=1}^{m}x_iV_{ji}<0,\end{aligned} \quad (3)$$

where $v_{ji}=F_{j+1,i}-F_{ji}$, $j=1,2,...,n-1$, and $i=1,2,...,m.$

In this case, the value ζ_j, $j=1,2,...,n-1$, characterizes the difference of the "usefulness" of the (j+1)-th and j-th objects.

The system of inequalities (3) with the addition of the positive variables x_{m+j}, $j=1,2,...,n-1$ is transformed into the system of equations

$$\sum_{i=1}^{m}x_iV_{ji}+x_{m+j}=0,\quad j=1,2,...,n-1. \quad (4)$$

Thus, the problem is reduced to the search for a solution of a homogeneous system of linear algebraic equations (4). This system always has the trivial solution:

$$x_1=x_2=...=x_n=x_{n+1}=...=x_{n+m-1}=0.$$

This solution is unique if the rank of the basic matrix of the system is equal to the number of variables. Otherwise, there are an infinite number of solutions. The peculiarity of the problem lies in the fact that specific requirements are imposed on the desired solution, i. e., the non-negativity of the variables $x_1,x_2,...,x_m$ and the positivity of the variables x_{m+j}, $j=1,2,...,n-1$. The number of additional variables can be reduced to one if the Chebyshev point is used as the solution of the system [24].

Now, taking into account the requirements for the signs of variables, the solution of the problem can be obtained using linear programming methods: thus, we find a set X that minimizes x_m+1 and satisfies the system of linear equations:

$$\sum_{i=1}^{m}x_iV_{ji}+x_{m+j}=0,$$

$$j=1,2,...,n-1. \quad (5)$$

In this case, in order to exclude the trivial solution

$$x_1=x_2=...=x_m=x_{m+1}=0,$$

system (5) requires adding one more equation:

$$\sum_{i=1}^{m}x_i=1, \quad (6)$$

which is the normalization condition for the coefficients of the regression equation. The resulting solution is used to calculate the "usefulness" of the objects according to formula (1).

A brief review of the known methods for solving problems of fuzzy regression analysis allows making the following conclusions.

1. In the problem with the given indistinctly explanatory and explained variables, there is no theoretically grounded criterion for assessing the quality of the results of regression analysis that would take into account possible significant differences in the accuracy of measurements of the explanatory and explained variables. This effect arises and manifests itself particularly negatively when, under conditions of a small sample of observations, no hypotheses regarding the laws of error distribution in measuring the variables can be reasonably accepted or rejected. This circumstance requires using fuzzy descriptions of the observed variables when solving practical problems of regression analysis the accuracy of which can be significantly different.

2. In the problem of comparator identification, the question of estimating the regression coefficients remains open for the case when the explanatory variables are not clearly defined.

The fundamental novelty of the problems arising in connection with this determines the importance of the research topic.

3. The aim and objectives of the study

The aim of the work is to improve the technology of a fuzzy regression analysis in the direction of developing a valid criterion for estimating the quality of regression and its use in solving practical problems.

In accordance with the stated goal, the main objectives are formulated as follows:

– to develop a reasonable criterion for assessing the quality of solving the problem of regression analysis under conditions where the explanatory and explained variables are not clearly defined;

– to devise a method for solving the problem of fuzzy regression analysis based on the selected criterion;

– to develop a method for solving the problem of comparator identification under conditions of fuzzy initial data.

4. The method for solving the problem of regression analysis under conditions of fuzziness of all initial data

Let us introduce the regression relation

$$x_0 + x_1F_1 + \ldots + x_nF_n = y, \tag{7}$$

where the explanatory variables $\{F_i\}$, $i = 1,2,\ldots,n$, and the explained variable y are fuzzy numbers with known membership functions.

Let N experiments be carried out, as a result of which the values of all the variables of the problem are obtained, $(N > n)$. Substitution of these values leads to the system of linear algebraic equations:

$$\begin{aligned} &x_0 + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n - a_{1,n+1} = 0,\\ &x_0 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n - a_{2,n+1} = 0,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &x_0 + a_{N1}x_1 + a_{N2}x_2 + \ldots + a_{Nn}x_n - a_{N,n+1} = 0. \end{aligned} \tag{8}$$

Here, for convenience, redesignations are made:

$$F_{ji} = a_{ji}, \quad y_j = a_{j,n+1},$$

$$j = 1,2,\ldots,N \text{ and } i = 1,2,\ldots,n.$$

We assume that the parameters of system (8) are Gaussian fuzzy numbers with the membership functions:

$$\mu(a_{ji}) = \exp\left\{-\frac{\left(a_{ji} - a_{ji}^{(0)}\right)^2}{2\sigma_{ji}^2}\right\}, \quad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,N. \tag{9}$$

Now we will say that relations (8) and (9) constitute a fuzzy system of linear algebraic equations. Let us solve this system [25, 26].

Let us introduce the set of fuzzy numbers:

$$\begin{aligned} &z_1 = \sum_{i=0}^{n} a_{1i}x_i - a_{1,n+1}, \; a_{10} = 1,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &z_N = \sum_{i=0}^{n} a_{Nj}x_i - a_{N,n+1}, \; a_{N0} = 1. \end{aligned} \tag{10}$$

We now introduce a clear system of linear algebraic equations, generated by relations (8) and (9), using the modal values $a_{ij}^{(0)}$ of the fuzzy numbers a_{ji},

$$i = 1,2,\ldots,N, \quad j = 0,1,\ldots,n+1:$$

$$\sum_{i=0}^{n} a_{ji}^{(0)}x_i - a_{i,n+1}^{(0)} = 0, \quad a_{0i}^{(0)} = 1, \quad i = 1,2,\ldots,N. \tag{11}$$

Let us introduce the following:

$$A^{(0)} = \left(a_{ji}\right), \quad X = \{x_i\}, \quad a_{j0} = 1,$$

$$i = 0,1,\ldots,n; \quad A_{n+1}^{(0)} = \left(a_{j,n+1}^{(0)}\right), \quad j = 1,2,\ldots,N.$$

Then the system of equations (11) in the matrix takes the form:

$$A^{(0)}X - A_{n+1}^{(0)} = 0. \tag{12}$$

The system of equations (12) is redefined and, possibly, incompatible. The natural solution is obtained by the method of least squares. We introduce the criterion of least squares:

$$J = \left(A^{(0)}X - A_{n+1}^{(0)}\right)^T \left(A^{(0)}X - A_{n+1}^{(0)}\right). \tag{13}$$

Minimizing (13) with respect to X, we obtain $X^0 = \{x_i^0\}$, $i = 0,1,\ldots,n$, as the solution of the linear equations system (LES) (11).

Then the clear solution of the fuzzy system of linear algebraic equations (8) and (9) will be the set $X=\{x_j\}$ $j = 0,1,\ldots,n$, minimizing the sum of the areas of the figures that are limited by the membership functions $\mu(z_j)$ of the fuzzy numbers $z_1, z_2, \ldots, z_N$, and being the least different from $X^{(0)}$ [26, 27]. The meaning of this criterion is understandable. Its use provides a set of clear numbers $x_0, x_1, \ldots, x_n$, for which the membership functions of the numbers $z_1, z_2, \ldots, z_N$ are the least blurred and have modal values that are as close to zero as possible.

Let us write the necessary relations, ensuring the solution of the system of equations (8) and (9) in the sense indicated above. In accordance with (9), we define the membership functions of the fuzzy numbers $z_1, z_2, \ldots, z_N$ given by (10). Thus,

$$\begin{aligned} \mu(z_j) &= \mu\left(\sum_{i=0}^{n} a_{ji}x_i - a_{i,n+1}\right) = \\ &= \exp\left\{-\frac{\left[z_j - \left(\sum_{i=0}^{n} a_{ji}^{(0)}x_i - a_{j,n+1}^{(0)}\right)\right]^2}{2\left(\sum_{i=0}^{n} \sigma_{ji}^2 x_i^2 + \sigma_{j,n+1}^2\right)}\right\} = \\ &= \exp\left\{-\frac{(z_j - m_j)^2}{2\sigma_j^2}\right\} \end{aligned}$$

and

$$m_j = \sum_{i=0}^{n} a_{ji}^{(0)}x_i - a_{j,n+1}^{(0)}, \quad \sigma_j^2 = \sum_{j=1}^{n} \sigma_{ji}^2 x_i^2 + a_{j,n+1}^2, \quad j = 1,2,\ldots,N.$$

Let $X^0 = \{x_i^0\}$, $i = 0,1,\ldots,n$ be the LES solution (11), whose parameters correspond to the modal values of the membership functions (12). In this case, since $m_z = m_x + m_y$ is the zero vector and the vector components $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ represent the discrepancies arising when the vector $z = x_1 - x_2$ is substituted in (8), then the natural measure of the deviation of the set X from the set $X^{(0)}$ is the sum of squares of the discrepancies, that is,

$$\begin{aligned} J_0 &= \left(A^{(0)}X - A_{n+1}^{(0)}\right)^{\mathrm{T}} \left(A^{(0)}X - A_{n+1}^{(0)}\right) = \\ &= \sum_{j=1}^{N}\left(\sum_{i=0}^{n} a_{ji}^{(0)}x_i - a_{j,n+1}^{(0)}\right)^2, \end{aligned}$$

where

$$A^{(0)} = \left(a_{ij}^{(0)}\right), \quad j = 1,2,\ldots,N, \quad i = 0,1,\ldots,n$$

and

$$A_{n+1}^{(0)} = \left(a_{1,n+1}^{(0)}, a_{2n+1}^{(0)}, \ldots, a_{N,n+1}^{(0)}\right)^{\mathrm{T}}, \quad X = \left(x_0 \;\; x_1 \;\; \ldots \;\; x_n\right)^{\mathrm{T}}.$$

In addition, we define the functional that determines the total area of the figures bounded by the membership functions $\mu\left(z_j\right)$ of the fuzzy numbers $z_1, z_2, ..., z_N$, as follows:

$$J_1 = \sum_{j=1}^{N} \int_{-\infty}^{\infty} \mu\left(z_j\right) dz_j.$$

Then the required clear solution of the system of equations (8) and (9) is obtained by minimizing the functional

$$J = J_1 + J_0 = \sum_{i=1}^{N} \left\{ \left[\int_{-\infty}^{\infty} \mu\left(z_i\right) dz_i \right] + \left(\sum_{j=1}^{n} a_{ij}^{(0)} x_j - a_{i,n+1}^{(0)} \right)^2 \right\}.$$

Since

$$\int_{-\infty}^{\infty} \mu\left(z_j\right) dz_j = \int_{-\infty}^{\infty} \exp\left\{ -\frac{(z_j - m_j)^2}{2\sigma_j^2} \right\} dz_i =$$

$$= \sqrt{2\pi}\sigma_j \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{ -\frac{(z_j - m_j)^2}{2\sigma_j^2} \right\} dz_i = \sqrt{2\pi}\sigma_j,$$

then

$$J = \\ = \sum_{j=1}^{N} \left\{ \left[\sqrt{2\pi} \left(\sum_{i=1}^{n} \sigma_{ji}^2 x_i^2 + \sigma_{j,n+1}^2 \right)^{\frac{1}{2}} \right] + \left(\sum_{i=0}^{n} a_{ji}^{(0)} x_i - a_{j,n+1}^{(0)} \right)^2 \right\}. \quad (14)$$

Furthermore,

$$\frac{dJ}{dx_k} = \sum_{j=1}^{N} \left[\frac{\sqrt{2\pi}\sigma_{kj}^2 x_k}{\left(\sum_{i=1}^{n} \sigma_{jk}^2 x_i^2 + \sigma_{j,n+1}^2 \right)^{\frac{1}{2}}} + 2\left(\sum_{i=1}^{n} a_{ji}^{(0)} x_i - a_{j,n+1}^{(0)} \right) a_{jk}^{(0)} \right] = 0,$$

$$k = 0, 1, ..., n.$$

The numerical solution of this nonlinear system of equations gives the required vector $X = \left(x_0 \quad x_1 \quad ... \quad x_n\right)$ of coefficients of the regression equation.

Finally, let us consider the problem of comparator identification for the case when the original data are not clearly described.

We will assume that the value of the i-th characteristic of the j-th object is a fuzzy number r_{ji} with the membership function

$$\mu_{ji}(r_{ji}), \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n.$$

Since the characteristics of the objects are fuzzy, the results of calculating the values of the "usefulness" level of the objects are also fuzzy:

$$Q_j(x) = \sum_{i=1}^{m} x_i r_{ji},$$

$$j = 1, 2, ..., n. \quad (15)$$

Let, for definiteness, r_{ji} be fuzzy numbers with a triangular membership function, that is,

$$\mu_{ij}(r_{ij}) = \begin{cases} 0, r_{ji} < a_{ji}, \\ \dfrac{r_{ji} - a_{ji}}{c_{ji} - a_{ji}}, a_{ji} \le r_{ji} < c_{ji}, \\ \dfrac{b_{ji} - r_{ji}}{b_{ji} - c_{ji}}, c_{ji} \le r_{ji} \le b_{ji}, \\ 0, r_{ji} > b_{ji}. \end{cases} \quad (16)$$

Since triangular fuzzy numbers are a particular case of numbers with a membership function of the (L–R) type, approximate calculations can be made by using the rules for performing operations accepted for fuzzy numbers of this type [26]. With this in mind, we define the membership functions of the fuzzy numbers $Q_i(X)$:

$$\mu_i(Q_j(X)) = \begin{cases} 0, & Q_j(X) < A_j, \\ \dfrac{Q_j(X) - A_j}{C_j - A_j}, & A_j \le Q_j(X) < C_j, \\ \dfrac{B_j - Q_j(X)}{B_j - C_j}, & C_j \le Q_j(X) < B_j, \\ 0, & Q_j(X) \ge B_j \end{cases} \quad (17)$$

and

$$A_j = \sum_{i=1}^{m} x_j a_{ji}, \quad C_j = \sum_{i=1}^{m} x_i c_{ji}, \quad B_j = \sum_{i=1}^{m} x_j b_{ji}.$$

If in the case under consideration, ranking (2) of the objects by the level of their "usefulness" is preserved, then the natural analogue (3) will be the fuzzy inequalities

$$\zeta_1(X) = Q_2(X) - Q_1(X) = \\ = x_1(r_{21} - r_{11}) + x_2(r_{22} - r_{12}) + ... + x_m(r_{2m} - r_{1m}) = \\ = \sum_{i=1}^{m} x_i w_{1i} < 0, \quad (18)$$

....

$$\zeta_{n-1}(X) = Q_n(X) - Q_{n-1}(X) = \sum_{i=1}^{m} x_i w_{n-1,i} < 0$$

and

$$w_{ji} = r_{j+1,i} - r_{ji}, \quad j = 1, 2, ..., n-1, \quad i = 1, 2, ..., m. \quad (19)$$

We pose the problem of finding the nonnegative set

$$X = (x_1, x_2, ..., x_n),$$

ensuring the fulfillment of inequalities (18).

The system of inequalities (18) with the addition of the positive variable x_{m+1} is transformed into the fuzzy system of linear algebraic equations

$$\sum_{i=1}^{m} x_i w_{ji} + x_{m+1} = 0, \quad j = 1, 2, ..., n-1, \quad (20)$$

where w_{ji} designates fuzzy numbers, the membership function of which is determined taking into account (16). Using the standard description of the membership function of triangular numbers of the L–R type in the form

$$r_{ji} = < c_{ji}, \alpha_{ji}, \beta_{ji} >_{L-R},$$

where c_{ji} is the mode of the number r_{ji}, $\alpha_{ji} = c_{ji} - a_{ji}$, and $\beta_{ji} = b_{ji} - c_{ji}$, as well as the rules for performing operations for fuzzy numbers of the $L-R$ type, we write the membership function of the number w_{ji}.

Then

$$w_{ji} = r_{j+1,i} - r_{ji} =< c_{j+1,i} - c_{ji}, \quad c_{ji} - a_{ji} + b_{j+1,i} - c_{j+1,i}$$

and

$$b_{ji} - c_{ji} + c_{j+1,i} - a_{j+1,i} \underset{LR}{>} =< m_{ji}, \hat{\alpha}_{ji}, \hat{\beta}_{ji} \underset{LR}{>}.$$

Here $\hat{m}_{ji}$ is the mode of the number w_{ji}, and $\hat{\alpha}_{ji}$ and $\hat{\beta}_{ji}$ are the left and right fuzziness coefficients.

Let us introduce the fuzzy numbers

$$Z_j(X) = \sum_{i=1}^{m} x_i w_{ji} + x_{m+1}, \quad j = 1,2,....,n-1$$

and write down the functions of their membership:

$$\mu(Z_j(X)) = \begin{cases} 0, & Z_j(X) < a_{1j} \\ \dfrac{Z_j(x) - a_1}{\sum_{i=1}^{m} x_i \alpha_{ji}}, & a_1 \le Z_j(X) < a_{2j}, \\ \dfrac{a_3 - Z_j(X)}{\sum_{i=1}^{m} x_i \beta_{ji}}, & a_2 \le Z_j(x) \le a_{3j}, \\ 0, & Z_i(x) > a_{3j}, \end{cases} \tag{21}$$

$$a_{1j} = \sum_{i=1}^{m} x_i (\hat{m}_{ji} - \hat{\alpha}_{ji}) + x_{m+1}, \quad a_{2j} = \sum_{i=1}^{m} x_i \hat{m}_{ji} + x_{m+1}$$

and

$$a_{3j} = \sum_{j=1}^{n} x_i (\hat{m}_{ji} + \hat{\beta}_{ji}) + x_{m+1}.$$

We transform the system of fuzzy equations $Z_j(X) = 0$, $j = 1,2,....,n-1$, that follows from (16) into an ordinary system of linear algebraic equations by specifying the fuzzy numbers w_{ji} as being equal to their modal values. In this case, we obtain

$$\sum_{i=1}^{m} x_i m_{ji} + x_{m+1} = 0, \quad j = 1,2,...,n-1. \tag{22}$$

In order to eliminate the trivial solution

$$x_i = 0, \quad i = 1,2,...,m, \quad x_i = 0, \quad i = 1,2,...,m,$$

of system (22), we add to it another normalizing equation (6). Let the set

$$X^{(0)} = (x_0^{(0)}, x_1^{(0)}, ..., x_m^{(0)}, x_{m+1}^{(0)})$$

be a solution of (6) and (22).

We use the definition of "a clear solution of a fuzzy system of linear algebraic equations" introduced in [26, 27]. In accordance with this definition, a clear solution of the system of equations (6) and (22) will be the set $X = (x_1, x_2, ..., x_{n+1})$, minimizing the sum of the areas of the figures bounded by the membership functions $\mu(Z_j)$ of the fuzzy numbers $Z_1, Z_2, ..., Z_{n-1}$ and the least deviating from $X^{(0)}$. The meaning of this definition is clear. Its use provides a set of the clear numbers $(x_1, x_2, ..., x_{n+1})$, as close as possible to the modal $X^{(0)}$, for which the membership functions of the fuzzy numbers $Z_1, Z_2, ..., Z_{m-1}$ are the least blurred. As a criterion for the compactness of the fuzzy-number membership functions Z_j, $j = 1,2,...,n-1$, the squares of the length of the intervals can be used as the carriers of the corresponding fuzzy numbers. Then the measure of the quality of solving the system of (6) and (22) will have the form:

$$J = \sum_{j=1}^{n-1} \left(\sum_{i=1}^{m} x_i (\beta_{ij} + \alpha_{ij}) \right)^2 + \sum_{i=1}^{m} (x_i - x_i^{(0)})^2. \tag{23}$$

Minimization (23) together with the normalization of condition (6) yields the desired set $X = (x_1, ..., x_{m+1})$.

5. Discussion of the findings

Thus, the study suggests methods for solving regression analysis problems under conditions of uncertainty in the values of explanatory and explained variables. For the case when all these variables are given by fuzzy numbers, a criterion for the quality of solving the problem is introduced and justified. The use of this criterion makes it possible to calculate the values of the regression experiments. The resulting set provides a compactness of the membership function of the fuzzy value predicted by the regression model of the explained variable and the proximity to the modal value. The required set is determined as a result of solving the fuzzy mathematical programming problem by the method proposed in [28]. The same technology was used to solve the problem of fuzzy comparator identification.

The possible continuation of this research is related to the difficulties of solving regression analysis problems with deeper uncertainties in the initial data. Let, for example, the values of variables be given in terms of fuzzy mathematics [29]. A possible approach to solving the problem in this case is the formation of fuzzy models of inaccurate descriptions of the variables [30]. Another variant of an inaccurately defined problem of regression analysis arises when, under conditions of a small sample of initial data, it is not possible to describe their uncertainty in terms of the probability theory or fuzzy mathematics. In this case, the natural way to solve the problem is to find the minimax solution under the assumption of the worst distribution densities (or membership functions), using the methods of continual linear programming [31].

6. Conclusions

1. A criterion for estimating the quality of solving the problem of fuzzy regression analysis is proposed and justified, taking into account possible significant differences in the accuracy of estimating explanatory and explained variables.

2. A method for solving the problem of fuzzy regression analysis is developed based on the selected criterion, using a fuzzy optimization technology.

3. A method for solving the problem of fuzzy comparator identification is developed when the results of measuring the explained variable are absent.

References

1. Linear Regression Analysis with Fuzzy Model [Text] // IEEE Transactions on Systems, Man, and Cybernetics. – 1982. – Vol. 12, Issue 6. – P. 903–907. doi: 10.1109/tsmc.1982.4308925
2. Yang, M.-S. Fuzzy least-squares algorithms for interactive fuzzy linear regression models [Text] / M.-S. Yang, H.-H. Liu // Fuzzy Sets and Systems. – 2003. – Vol. 135, Issue 2. – P. 305–316. doi: 10.1016/s0165-0114(02)00123-9
3. Chen, F. Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients [Text] / F. Chen, Y. Chen, J. Zhou, Y. Liu // Engineering Applications of Artificial Intelligence. – 2016. – Vol. 47. – P. 16–24. doi: 10.1016/j.engappai.2015.02.011
4. Ishibuchi, H. Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks [Text] / H. Ishibuchi, M. Nii // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 273–290. doi: 10.1016/s0165-0114(98)00370-4
5. Chang, Y.-H. O. Fuzzy regression methods – a comparative assessment [Text] / Y.-H. O. Chang, B. M. Ayyub // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 187–203. doi: 10.1016/s0165-0114(99)00091-3
6. Yen, K. K. A linear regression model using triangular fuzzy number coefficients [Text] / K. K. Yen, S. Ghoshray, G. Roig // Fuzzy Sets and Systems. – 1999. – Vol. 106, Issue 2. – P. 167–177. doi: 10.1016/s0165-0114(97)00269-8
7. Charfeddine, S. Fuzzy linear regression: application to the estimation of air transport demand [Text] / S. Charfeddine, F. Antonio C. Mora-Camino, M. De Coligny // FSSCEF 2004, International Conference on Fuzzy Sets and Soft Computing in Economics and Finance. – 2004. – P. 350–359. – Available at: https://hal-enac.archives-ouvertes.fr/hal-01022443
8. Zaychenko, Yu. P. Nechetkiy metod gruppovogo ucheta argumentov pri neopredelennyh iskhodnyh dannyh [Text] / Yu. P. Zaychenko // Systemni doslidzhennia ta informatsiini tekhnolohii. – 2007. – Issue 3. – P. 100–112.
9. Zgurovskiy, M. Modeli i metody prinyatiya resheniy v nechetkih usloviyah [Text] / M. Zgurovskiy, Yu. Zaychenko. – Kyiv: Naukova dumka, 2013. – 275 p.
10. Celmins, A. Least squares model fitting to fuzzy vector data [Text] / A. Celmins // Fuzzy Sets and Systems. – 1987. – Vol. 22, Issue 3. – P. 245–269. doi: 10.1016/0165-0114(87)90070-4
11. Muzzioli, S. A comparison of fuzzy regression methods for the estimation of the implied volatility smile function [Text] / S. Muzzioli, A. Ruggieri, B. De Baets // Fuzzy Sets and Systems. – 2015. – Vol. 266. – P. 131–143. doi: 10.1016/j.fss.2014.11.015
12. Diamond, P. Fuzzy least squares [Text] / P. Diamond // Information Sciences. – 1988. – Vol. 46, Issue 3. – P. 141–157. doi: 10.1016/0020-0255(88)90047-3
13. Chang, Y.-H. O. Hybrid fuzzy least-squares regression analysis and its reliability measures [Text] / Y.-H. O. Chang // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 225–246. doi: 10.1016/s0165-0114(99)00092-5
14. Yang, M.-S. Fuzzy least-squares linear regression analysis for fuzzy input–output data [Text] / M.-S. Yang, T.-S. Lin // Fuzzy Sets and Systems. – 2002. – Vol. 126, Issue 3. – P. 389–399. doi: 10.1016/s0165-0114(01)00066-5
15. Kao, C. A fuzzy linear regression model with better explanatory power [Text] / C. Kao, C.-L. Chyu // Fuzzy Sets and Systems. – 2002. – Vol. 126, Issue 3. – P. 401–409. doi: 10.1016/s0165-0114(01)00069-0
16. Shtovba, S. D. Nechetkaya identifikaciya na osnove regressionnyh modeley parametricheskoy funkcii prinadlezhnosti [Text] / S. D. Shtovba // Problemy upravleniya i informatiki. – 2006. – Issue 6. – P. 1–8.
17. Chachi, J. Multiple fuzzy regression model for fuzzy input-output data [Text] / J. Chachi, S. M. Taheri // Iranian Journal of Fuzzy Systems. – 2016. – Vol. 13, Issue 4. – P. 63–78.
18. Zak, Yu. Fuzzy-regressionnye modeli v usloviyah nalichiya v statisticheskiy vyborke nechislovoy informacii [Text] / Yu. Zak // System Research & Information Technologies. – 2017. – Issue 1. – P. 88–96.
19. Jung, H.-Y. Fuzzy linear regression using rank transform method [Text] / H.-Y. Jung, J. H. Yoon, S. H. Choi // Fuzzy Sets and Systems. – 2015. – Vol. 274. – P. 97–108. doi: 10.1016/j.fss.2014.11.004
20. Pushpa, B. A Least Absolute Approach to Multiple Fuzzy Regression Using Tw- Norm Based Operations [Text] / B. Pushpa, R. Vasuki // International Journal of Fuzzy Logic Systems. – 2013. – Vol. 3, Issue 2. – P. 73–84. doi: 10.5121/ijfls.2013.3206
21. Namdari, M. Fuzzy logistic regression with least absolute deviations estimators [Text] / M. Namdari, J. H. Yoon, A. Abadi, S. M. Taheri, S. H. Choi // Soft Computing. – 2014. – Vol. 19, Issue 4. – P. 909–917. doi: 10.1007/s00500-014-1418-2
22. Ubale, A. B. Fuzzy Regression Model and Its Application: A Review [Text] / A. B. Ubale, S. L. Sananse // International Journal of Innovative Research in Science, Engineering and Technology. – 2015. – Vol. 4, Issue 11. – P. 10853–10860.
23. Kryuchkovskiy, V. V. Vedenie v normativnuyu teoriyu prinyatiya resheniy [Text] / V. V. Kryuchkovskiy, Eh. G. Petrov, N. A. Sokolova, V. E. Hodakov. – Herson: Grin' D. S., 2013. – 284 p.
24. Zuhovickiy, S. I. Lineynoe i vypukloe programmirovanie [Text] / S. I. Zuhovickiy, L. I. Avdeeva. – Moscow: «Nauka», 1967. – 460 p.
25. Sira, O. V. A New Approach for Resolving Equations with Fuzzy Parameters [Text] / O. V. Sira, K. H. Al-Shqeerat // European Journal of Scientific Research. – 2009. – Vol. 38, Issue 4. – P. 619–625.
26. Seraya, O. V. Linear Regression Analysis of a Small Sample of Fuzzy Input Data [Text] / O. V. Seraya, D. A. Demin // Journal of Automation and Information Sciences. – 2012. – Vol. 44, Issue 7. – P. 34–48. doi: 10.1615/jautomatinfscien.v44.i7.40
27. Raskin, L. G. Nechetkaya matematika [Text] / L. G. Raskin, O. V. Seraya. – Kharkiv: Parus, 2008. – 352 p.
28. Raskin, L. Method of solving fuzzy problems of mathematical programming [Text] / L. Raskin, O. Sira // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 5, Issue 4 (83). – P. 23–28. doi: 10.15587/1729-4061.2016.81292

29. Pawlak, Z. Rough sets [Text] / Z. Pawlak // International Journal of Computer & Information Sciences. – 1982. – Vol. 11, Issue 5. – P. 341–356. doi: 10.1007/bf01001956
30. Raskin, L. Fuzzy models of rough mathematics [Text] / L. Raskin, O. Sira // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 6, Issue 4 (84). – P. 53–60. doi: 10.15587/1729-4061.2016.86739
31. Raskin, L. G. Prikladnoe kontinual'noe lineynoe programmirovanie [Text] / L. G. Raskin, I. O. Kirichenko, O. V. Seraya. – Kharkiv, 2013. – 293 p.

UDC 519.6
DOI: 10.15587/1729-4061.2017.108327

CONSTRUCTION OF INTERPOLATION METHOD FOR NUMERICAL SOLUTION OF THE CAUCHY'S PROBLEM

R. Bihun
Postgraduate student*
E-mail: bigunroman@ukr.net
G. Tsehelyk
Doctor of Physical and Mathematical Sciences, Professor, Head of Department*
E-mail: kafmmsep@lnu.edu.ua
*Department of Mathematical modeling of socio-economic processes
Ivan Franko National University of Lviv
Universytetska str., 1, Lviv, Ukraine, 79000

Побудовано інтерполяційний чисельний метод розв'язування задачі Коші для звичайних диференціальних рівнянь першого порядку за допомогою апарату некласичних мінорант та діаграм Ньютона функцій, заданих таблично. Цей метод дає точніші результати від методу Ейлера у випадку опуклої функції. Доведено обчислювальну стійкість методу, тобто похибка початкових даних не нагромаджується. Також показано, що метод має другий порядок точності

Ключові слова: міноранта Ньютона, диференціальні рівняння, задача Коші, діаграма Ньютон, опукла функція

Построен интерполяционный численный метод решения задачи Коши для обыкновенных дифференциальных уравнений первого порядка с помощью аппарата неклассических минорант и диаграмм Ньютона функций, заданных таблично. Этот метод дает более точные результаты по сравнению с методом Эйлера в случае выпуклой функции. Доказана вычислительная устойчивость метода, то есть погрешность начальных данных не накапливается. Также показано, что метод имеет второй порядок точности

Ключевые слова: минорант Ньютона, дифференциальные уравнения, задача Коши, диаграмма, выпуклая функция

1. Introduction

Cauchy's problem is one of the main problems in the theory of differential equations, which comes down to finding a solution (integral) of the differential equation that satisfies initial conditions (original data).

Over many years, a numerical solution of the Cauchy's problem has been the focus of attention by scientists as it is widely used in different areas of science and technology. That is why there are a large number of developed methods for it. In spite of this, however, new methods are being devised, some of them with better properties than those preceding.

Cauchy's problem usually emerges during analysis of the processes predetermined by the differential law and original state. Mathematical notation of such equations is an equation and the initial condition.

The difference between the boundary-value problems and the Cauchy's problem is that the region over which the desired solution should be determined is not specified in the latter in advance. However, the Cauchy's problem can be considered as one of the boundary-value problems.

2. Literature review and problem statement

Numerical methods of the Cauchy's problem solution are divided into 3 groups [1]:

– one-point;
– multipoint (methods of prediction and correction);
– methods with automatic choice of integration step.

The one-point methods include methods that have certain common features, such as:

1. Underlying all one-point methods is the function decomposition into Taylor's series, which preserves members that have h in a power to k inclusive. An integer k is called the order of the method. Error on a step has an order of $k+1$.

2. All one-point methods do not require a valid computation of derivatives, because only the function itself is calculated, however, one may require its values in some intermediate points. This entails, of course, additional cost of time and effort.

3. In order to receive information in a new point, it is necessary to have data only from the previous point. This property can be called "self-starting". A capability to "self-start" makes it possible to easily change the magnitude of step h.