UDC 517.957

S.B. Prykhodko, Dr. of Sc., Prof., N.V. Prykhodko, Phd, Assist. Prof., K.S. Pugachenko

MODELING A DEPENDENT VARIABLE OF NON-LINEAR REGRESSION ON THE BASIS OF NORMALIZING TRANSFORMATIONS

Sometimes in modeling a variety of technical, organizational and economic systems there is a need to predict the value of dependent variable of non-linear regression model from knowledge of the values of one or more independent variables. Herewith the dependent variable is a non-Gaussian random variable.

It is known, any regression model is a dependency of the random dependent variable on the sum of conditional expectation (non-random component) and residuals (random component). Both components depend on the independent variables. For non-linear regression model the conditional expectation is non-linear function of one or more independent variables and the random component is a non-Gaussian random variable. Usually the acceptance-rejection method is used for simulation of continuous random variables from knowledge of the non-Gaussian joint distribution in the form of the probability density function. This is the main problem in modeling a dependent variable of non-linear regression.

We propose the novel technique for modeling a dependent variable of nonlinear regression without the non-Gaussian joint probability density function. The technique is based on bijective multivariate normalizing transformations. At the beginning of the non-linear regression equation and its prediction intervals must be constructed on the basis of the multivariate normalizing transformation [1]. After that the technique is performed in five steps: first, prediction non-linear regression equation result and the values of boundaries of prediction interval of non-linear regression for the values of one or more independent variables are defined; second, the normalized values are calculated on the basis of the normalizing transformation; third, a value of the Gaussian random variable for the normalized values of boundaries of prediction interval (with zero mean and variance equal to one for Johnson translation system [1]) is simulated by the acceptance-rejection method; then, we calculate the value of normalized dependent variable of the linear regression as the sum of the normalized value of prediction non-linear regression equation result and the simulated value of Gaussian random variable; finally, the value of dependent variable of the non-linear regression is defined by applying the inverse transformation to the value of normalized dependent variable.

It should be noted, for simulation of a value of the Gaussian random variable we can use another method, for example the method based on Johnson normalizing transformation (Johnson translation system) for S_U family [2]. This method, unlike the acceptance-rejection method, uses only one uniform random number to generate a value of the Gaussian random variable. The boundaries of prediction interval of non-linear regression of one independent variable are defined by the technique from [3].

The results of modeling and simulating two dependent variables of the non-linear regressions on the basis of Johnson normalizing transformation for S_B family are presented. First non-linear regression model for estimating software testing effort contains four independent variables – Use Cases, Tester Year of experience, Developer Year of experience, Test Cases, second non-linear regression model for software defect prediction has seven independent variables: Requirement Errors, Coding Errors, KLOC (number of thousands of lines of code), Requirement Pages, Design Pages, Total Test Cases and Total Effort Days.

List of references:

1. Prykhodko S.B. Developing the software defect prediction models using regression analysis based on normalizing transformations / S. B. Prykhodko // Сучасні проблеми тестування прикладного програмного забезпечення: збірник тез доповідей науково-практичного семінару, Полтава, 25-26 травня 2016 р. – Полтава, 2016. – С.6-7.

2. Приходько С.Б. Моделювання гаусівських випадкових величин із використанням перетворення Джонсона із сім'ї S_U / С. Б. Приходько // Інформатика та математичні методи в моделюванні. – 2015. – т.5, № 1. – С.92-97.

3. Prykhodko S.B. Statistical anomaly detection techniques based on normalizing transformations for non-Gaussian data / S. B. Prykhodko // Computational Intelligence (Results, Problems and Perspectives): Proceedings of the International Conference, May 12-15, 2015, Kyiv-Cherkasy, Ukraine / Vitaliy Ye. Snytyuk (Editor). – Cherkasy: editor July Chabanenko, 2015. – p.286-287.