Approaches to the construction of nonlinear models in fuzzy environment

D T Muhamediyeva, J Sayfiyev
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

Abstract. The Bayesian methods to the problems of statistical estimation when building nonlinear models are considered. Bayesian methods can be used to construct linear and nonlinear regression models with non-Gaussian laws for the distribution of probabilities of random observation errors. We consider continuous and discrete processes that can be described by statistical models. For this purpose, a description of computational Bayesian procedures and recommendations for the construction of nonlinear regression models are given.

1. Introduction
One of the main problems of statistical modelling is the construction of models that allow to identify causal relationships between variables. The solution to this problem is quite difficult. The discrepancy between the simulation results and the ratios that actually occur is due to a number of reasons, in particular, due to violations of the basic premises of the regression analysis. They are most often violated due to multicollinearity [1–2]. This phenomenon leads to estimates of parameters with large variance, which in many cases does not allow them to be meaningfully interpreted, for example, because of incorrect signs. Obviously, the use of a priori information is a powerful tool for constructing the above-mentioned type of regression models. In particular, it reduces the effect of multicollinearity on parameter estimates. However, the formalization of a priori information due to its statistical uncertainty is not always a simple matter [3–7]. One of the possible approaches to solving this problem is the use of fuzzy mathematics methods.
If we are dealing with uncertainty, when we do not have enough observations to correctly confirm this or that distribution law or we observe objects that, strictly speaking, cannot be called homogeneous, then there is no classical statistical sample. At the same time, we, even without having a sufficient number of observations, tend to imply that behind them is a manifestation of some law. We cannot estimate the parameters of this law quite accurately, but we can come to a definite agreement on the form of this law and on the range of variation of key parameters included in its mathematical description. To solve such problems, statistical models have been developed using the theory of fuzzy sets [8].
A description of computational Bayesian procedures and recommendations for constructing nonlinear regression models with fuzzy initial information have been developed. We consider continuous and discrete processes that can be described by statistical models using the theory of fuzzy sets.

2. Statement of the task
Statement of the problem of construction mathematical model include:
on the basis of a priori information of the form of the functional dependence between the output \( Y \) and the input \( \mathbf{x} \) by variables:

\[
Y = \varphi_j(\mathbf{a}, \mathbf{x}), \quad j = 1, \ldots, m,
\]

where \( \mathbf{a} \) – k-dimensional vector of unknown model parameters
m – number of valid mathematical models that can be used to describe the process;
determination of the type of observation equation describing the nature of the effect of random errors of observation on the measurement result of controlled quantities:

\[
y = h_y(Y, \varepsilon_y), \quad \mathbf{x}_i = h_{\mathbf{x}, i}(\mathbf{x}, \varepsilon_{\mathbf{x}, i}),
\]

where \( \varepsilon_y \) – random observation error of output value;
\( \varepsilon_{\mathbf{x}, i} \) – random observation error of I-th input value;
h – data characterization;
n – the number of controlled input values; if nothing is known about the nature of the effect of a random observation error, it is assumed to be additive:

\[
y = Y + \varepsilon_y, \quad \mathbf{x}_i = \mathbf{x}_i + \varepsilon_{\mathbf{x}, i};
\]
determination of the probability distribution of the observation error \( f(\varepsilon) \) based on the results of a preliminary experimental study of the data.
In the absence of a priori information on the probability distribution, the observation error is assumed to be Gaussian, the probability density of which

\[
f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{\varepsilon^2}{2\sigma^2} \right\},
\]

where \( \sigma^2 \) – variance of observation error.
For nonstationary processes, the functional time dependence of the regression model parameters is taken into account [4,5].
Dynamic objects are described by mathematical models in the form of difference equations.
Discrimination (distinction) of models is carried out in order to select the best model from a given set of models by paired comparison by the criterion of likelihood ratio, in particular:
– sample values of discrimination statistics and its first four points are calculated;
– there are thresholds of discrimination;
– the conditions for the end of the discrimination procedure are checked.

### 3. Finding thresholds of discrimination

Discriminating threshold called bound of the interval \((z_1, z_2)\), inside which with probability \( P = 1 - \alpha \),
where \( \alpha \) – the specified level of significance a value of zero \( \mu_t \), are the values of the statistic

\[
z = \sum_{i=1}^{N} \frac{\lambda_i - N\mu_t}{\sqrt{N\mu_t^2}},
\]

where \( \mu_t \) – average value
\( \mu_t \) – variance of the random variable \( \lambda \),
N – number of observations.
Probability distribution function of probability \( z \) with accuracy to terms of order \( N^{-1.5} \) is approximated by the expression

\[
F(z) = \Phi(z) - \frac{1}{\sqrt{N}} \left( \frac{\alpha_2}{3!} \Phi^{(3)}(z) + \frac{1}{N} \frac{1}{4!} \alpha_4 \Phi^{(4)}(z) + \frac{10}{6!} \alpha_6 \Phi^{(6)}(z) \right);
\]

here \( \Phi(z), \Phi^{(i)}(z) \) – normal probability distribution function with zero mean and unit variance and its derivatives;
\( \alpha_3 \) and \( \alpha_4 \) – skewness and kurtosis values \( \lambda \), calculated by the formulas:
\[
\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}}; \quad \alpha_4 = \frac{\mu_4}{\mu_2^2} - 3.
\]

Thresholds of discrimination \( z_1 \) and \( z_2 \) are found by numerical solution of nonlinear equations:
\[
F(z_1) = 0.5 \alpha; \quad F(z_2) = 1 - 0.5 \alpha,
\]
where \( \alpha \) – specified level of significance \((0.01, 0.1)\).

To solve the equations are taken acceptable boundary values \( z_{\min} \) and \( z_{\max} \), and then use stepper motors on the axis \( z \) first from \( z_{\min} \), then from \( z_{\max} \) following values are found \( z_1 \) and \( z_2 \), which with a given accuracy are solutions to the equations.

Procedure of discrimination is over if the value computed by the formula
\[
z = \frac{\sum_{i=1}^{N} \lambda_i}{\sqrt{N \mu_2}}
\]
satisfies one of conditions: \( z > z_2 \); \( z < z_1 \).

When the first condition is accepted by the first model when the second – second model. Procedure discrimination, if the disparity \( z_1 \leq z \leq z_2 \). Under this condition, the models are considered indistinguishable (equivalent) for a given number of observations. In this case, it is possible to adopt any of the compared models, for example, the simplest one, or to continue the process of observation and discrimination.

4. Calculation point of the empirical Bayesian estimates of the parameters of the regression model

Empirical Bayesian estimation is used for estimating the model parameters of the observed process based on the accumulated experimental data on other similar processes belonging to the same class, the a priori probability density of estimated parameters for empirical Bayesian estimation is according to the results of observation of several processes.

Empirical Bayesian estimation of parameters of regression models the \( P \)-th observed process with quadratic loss function is calculated by the formula
\[
\hat{\Theta}_P = \frac{\sum_{j=1}^{T} \sum_{k=1}^{M} \bar{\Theta}(S) l(Y_j / \bar{\Theta}(S)) l(Y_j / \bar{\Theta}(S)) \pi(\bar{\Theta}(S)) \left[ \sum_{s=1}^{M} l(Y_j / \bar{\Theta}(S)) \pi(\bar{\Theta}(S)) \right]^{-1}}{\sum_{j=1}^{T} \sum_{k=1}^{M} l(Y_j / \bar{\Theta}(S)) l(Y_j / \bar{\Theta}(S)) \pi(\bar{\Theta}(S)) \left[ \sum_{s=1}^{M} l(Y_j / \bar{\Theta}(S)) \pi(\bar{\Theta}(S)) \right]^{-1}},
\]
where \( Y_j = \{y_{j_1}, y_{j_2}, \ldots, y_{j_{N_j}}\} \)
\[
Y_j = \frac{\sum_{k=1}^{N} Y_j^k \mu_j / \sum_{k=1}^{N} \mu_j}{\sum_{k=1}^{N} \mu_j / \sum_{k=1}^{N} \mu_j}
\]
– experimental data for the \( j \)-th technological process;
\[
\pi(\Theta) \] – subjective a priori probability density of the parameters;
\[
l(Y / \bar{\Theta}) \] – a likelihood function;
\[
N_j \] – number of observations;
\[
T \] – number of technological processes, similar to the observed \( p \)-s process;
\[
M \] – number of statistical tests in the procedure of the method of Monte Carlo;
\[
(\bar{\Theta})[S] \] – vector of pseudorandom numbers uniformly distributed in the domain \( \Omega_\theta \).
If the a priori probability density of the regression model parameters is unknown, then \( \pi(\Theta) \) selected as a uniform density or as a normal density.

Likelihood function for an independent normal additive observation error is written as

\[
l(Y / \Theta) = \frac{1}{(\sigma^2 \sqrt{2\pi})^N} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \phi[\bar{a}, \bar{x}])^2 \right\},
\]

where \( N \) – number of observations;
\( \sigma^2 \) – the variance of the observation error;
\( \phi(\bar{a}, \bar{x}) \) – regression function;
\( \bar{a} \) – regression parameters;
\( \Theta = [\bar{a}, \sigma^2] \) – estimated parameters.

Calculation of the subjective Bayesian point estimates. Subjective Bayesian estimates are used to estimate the model parameters of individual processes. The a priori probability density of the parameters is chosen as a weight function reflecting our subjective knowledge of the probabilities of the values of the estimated parameter [6].

Subjective point Bayesian estimation of parameters of regression model of technological process at a quadratic function of losses is calculated by the formula

\[
\hat{\Theta} = \frac{\sum_{s=1}^{M} \Theta[S] \cdot l(Y / \Theta[S]) \pi(\Theta[S])}{\sum_{s=1}^{M} l(Y / \Theta[S]) \pi(\Theta[S])},
\]

where \( l(Y / \Theta) \) – likelihood function;
\( \pi(\Theta) \) – subjective a priori probability density of the estimated parameters;
\( M \) – number of statistical tests in the Monte Carlo procedure;
\( \Theta[S] \) – vector of pseudorandom numbers uniformly distributed in a given region \( \Omega_\Theta \).

Accuracy of calculations is controlled by the value of the coefficient of variation

\[
\hat{\delta} = \frac{1}{I} \sqrt{\frac{D}{M}}
\]

and evaluating the integral

\[
I = \int_{\Omega} l(Y / \Theta) \pi(\Theta) d\Theta,
\]

calculated according to the formula

\[
\hat{I}[j] = \frac{1}{j} \sum_{s=1}^{j} l(Y / \Theta[S]) \pi(\Theta[S]) .
\]

Dispersion \( \hat{D} \) is found by the following recurrent formula:

\[
\hat{D}[j] = \frac{j-1}{j} (\hat{D}[j-1] + \frac{1}{j} (l(Y / \Theta[j]) \pi(\Theta[j]) - \hat{I}[j-1]^2), \quad j = 1, 2, ..., M,
\]

where \( M \) – specified number of Monte Carlo statistical tests \( M \approx 10^3 \div 10^4 \).

Procedure for calculating estimates is terminated in the following cases: if the number of statistical tests \( j \) exceeds the specified number \( M \), or if the value of \( \hat{\delta} \) it turns out to be less than the given relative accuracy \( \delta_0 = 0,1 \div 0,001 \).

Thus, in the Bayesian approach to the problems of statistical estimation with fuzzy initial information, the parameter estimates are found from the minimum condition of the risk function [7]:

\[
R(\Theta) = \int_{\Omega} W(\hat{\Theta}(Y) - \Theta) dF(Y) dG(\Theta),
\]

where \( W \) – loss function;
\( F(Y) \) – probability distribution function of observations;
5. Computational procedure of Bayesian nonlinear estimation

The Bayesian nonlinear estimation computational procedure consists of a large number of simple operations. Discrimination (discrimination) of models is carried out in order to select the best model from a given set of models by pairwise comparison by the likelihood ratio criterion.

For discrimination of models, the value calculated by the formula

\[ \log \left( \frac{f_1(y / \Theta_j)}{f_2(y / \Theta_j)} \right) \]

where \( f_j(y / \Theta_j) \) – probability density of \( Y \) for the \( j \)-th model \((j=1,2)\).

\( \Theta_j \) – model parameter vector including regression function parameters \( \phi(\vec{a}, \vec{x}) \) and parameters of the probability distribution function of observation errors.

Moments \( \nu \)-th order of magnitude \( \lambda \) calculated by the Monte Carlo method by the formula:

\[ m_\nu = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{l} \log \left( \frac{f_1(y_i / \Theta_j(l))}{f_2(y_i / \Theta_j(l))} \right) \pi_j(\Theta_j) \pi_\nu(\Theta_\nu), \]

where \( N \) is the number of observations;
\( \gamma \) – number of statistical tests per observation \((\gamma = 1 \pm 10)\);
\( \Theta_j(l) - K_j \)-dimensional vector of pseudo-random numbers with a uniform probability distribution in a rectangular domain \( \Omega_\theta \) with the sides \((c_i, d_i), (i = 1, 2, ..., K_j)\);
\( \pi_j(\Theta_j) \) – prior probability density parameters of the \( j \)-th model. Pseudorandom numbers are generated by a standard program on a computer.

Central moments of the magnitude are calculated by the formula

\[ \mu_\nu = m_\nu - m_\nu^2; \]
\[ \mu_4 = m_4 - 3m_2m_2 + 2m_4^2; \]
\[ \mu_4 = m_4 - 4m_2m_2 + 6m_4^2m_2 - 3m_4^4. \]

If the prior density of the model parameters is unknown, then \( \pi_j(\Theta_j) \) selected or in the form of a uniform probability density in the domain \( \Omega_\theta \):

\[ \pi_j(\Theta_j) = \prod_{i=1}^{K_j} \frac{1}{(d_i - c_i)} \text{at } \Theta_j = (c_i, d_i), \]

where \( \Theta_0 \) - \( i \)-th component of \( j \)-th \( K_j \)-dimensional vector parameter \( \Theta_j \) or as normal probability density

\[ \pi_j(\Theta_j) = \prod_{i=1}^{K_j} \frac{1}{\sigma_{ij} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_{ij}^2}(\Theta_j - \Theta_0)^2 \right\}, \]

where \( \Theta_{ji} = \{\Theta_{j,1}, \Theta_{j,2}, ..., \Theta_{j,K}\} \) - center point of the domain \( \Omega_\theta \);
\( \sigma_{ij} \) – standard deviation approximately equal to the range of variation of the \( i \)-th component of the \( j \)-th parameter \( \Theta_j \).

It should be noted that with increasing number of observations, the choice of subjective a priori probability density \( \pi(\Theta) \) little effect on the results of statistical processing of experimental data.

6. Making calculations

As the object of study is considered the yield of cotton. When it is possible to predict weather-forming factors with a long lead time, then the following factors are determined that affect cotton yields \([58] \):

\( x_i \) – cotton crop area;
\[ y = A_i x_1 a_i x_2 a_i x_3 a_i x_4 a_i x_5 a_i x_6 a_i x_7 a_i, \]  

where \( y \) – cotton yield;  
\( A_i \) – coefficient indicating the effect of unaccounted factors;  
\( a_i, a_2, \ldots, a_7 \) – coefficients indicating the proportion of influence factors.

When the dependence of yield on weather-forming factors and on the length of the vegetation period, which depend on solar activity, is known, a mathematical model of yield can also be constructed:

\[ y = a_0 + a_y b + (a_{i0} + a_{i1} \sin \frac{2\pi}{50}(t - 6) + \sum_{\rho=2}^{2} A_{\rho} \sin \frac{2\pi}{90 - M_{\rho}}(t - K_{\rho})). \]  

It is assumed that the error in observing the output quantity is an independent additive and normally distributed fuzzy quantity.

The likelihood function in this case is written as

\[ I(Y / \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right\}, \]

where \( N \) – number of independent observations  
\( \sigma^2 \) – variance of observation error  
\( \hat{y}_i \) – the value of the regression function calculated by equation (1) or (2) for the \( i \)-th moment of time.

To select the best of the two cotton yield models described by equations (1), (2), the models were discriminated according to the experimental cotton yield curves. A preliminary assessment of the parameters and analysis of the likelihood function showed that the prior probability density parameters of the compared models can be chosen uniform over the prior intervals \((a_i; a_i')\). Here \( a_i \) – described in the form of fuzzy subsets, i.e. the member ship functions of the corresponding subsets are given:

\[ a_i = \left\{ \mu_{a_i}, (a_i; a_i') \right\}, \]

where

\[ \mu_{a_i} = e^{-112500(x-0.77)^2}, \]
\[ (a_i; a_i') = (0.75;0.79), \]
\[ \mu_{a_2} = e^{-1125000(x-0.097)^2}, \]
\[ (a_i; a_i') = (0.095;0.099), \]
\[ \mu_{a_3} = e^{-20000(x-0.365)^2}, \]
\[ (a_i; a_i') = (0.350;0.380), \]
\[ \mu_{a_4} = e^{-11250(x+0.76)^2}, \]
\[ (a_i; a_i') = (-0.78;-0.74), \]
\[ \mu_{a_5} = e^{-11250(x+0.42)^2}, \]
\[ (a_i; a_i') = (0.40;0.44), \]
\[ \mu_{a_6} = e^{-11250(x+0.22)^2}, \]
\[ (a_i; a_i') = (0.20;0.24), \]
\[ \mu_{a_7} = e^{-45000000(x+0.002)^2}, \]
\[ (a_i; a_i') = (-0.003;-0.001), \]
\[ \mu_{a_8} = e^{\frac{1800(x-1.685)^2}{289}}, \]
\[ (a_i; a_i') = (1.60;1.77), \]
\[ \mu_{a_9} = e^{-450000(x-0.998)^2}, \]
\[ (a_i; a_i') = (0.997;0.999), \]
The discrimination procedure ends if the value calculated by the formula
\[ z = \sqrt{\frac{\sum \lambda_i}{N}} \]
satisfies one of the conditions: \( z > z_2 \), \( z < z_1 \).
When the first condition is fulfilled, the first model is adopted, when the second condition is fulfilled - the second model.
The discrimination procedure continues if there is inequality \( z_1 \leq z \leq z_2 \).

7. Conclusion
Thus, the advantage of the procedure for calculating Bayesian estimates is its convergence for an arbitrary type of non-linear, unambiguous regression models, which is due to the convergence of the procedure of the Monte Carlo method used to calculate multidimensional integrals. The use of Bayesian methods improves the accuracy of estimation due to the use of a priori information.
Processing of the calculated data was carried out under the program "Discrimination of models". As shown by the results of discrimination from several experimental data, cotton yields are more accurately described by the second model (2).

References
[1] Spokoiny V 2013 Basics of Modern Parametric Statistics (Springer) p 239
[2] Lazaro-Gredilla M Titsias M 2011 Variational heteroscedastic Gaussian process regression ICML
[3] MacKay DJ C 2005 Information theory, inference and learning algorithms (Cambridge university press) p 640
[4] Ghosh JK, Mohan D and Tapas S 2006 An Introduction to Bayesian analysis (New York: Springer) p 366
[5] Ghosh J K, Ramamoorthi R V 2003 Bayesian nonparametrics (Springer) p 324
[6] Kleijn B A van der Vaart H van Zanten 2013 Lectures on Nonparametric Bayesian Statistics (Springer)
[7] Rasmussen C E and Williams C K 2006 Gaussian processes for machine learning (Cambridge-London: The MIT Press) p 266
[8] Zadeh L A 2002 Journal of Statistical Planning and Inference p 105