关于贝叶斯多维矩阵多项式的经验回归
ON THE BAYESIAN MULTIDIMENSIONAL-MATRIX POLYNOMIAL EMPIRICAL REGRESSION

Vladimir S. Mukha
Doctor of Technical Sciences, Professor
Irina G. Malikova
Senior Lecturer
Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

抽象。 提出并解决了输入变量回归函数中多项式参数估计的问题。 回归函数的输入和输出变量是多维矩阵。 假设回归函数的参数是具有高斯分布，已知平均值和离散矩阵的随机独立多维矩阵。 解决该问题的方法是在多维矩阵未知数中使用线性代数方程的多维矩阵系统-函数回归参数。 我们考虑了二次回归函数的特殊情况，为此我们获得了用于参数计算的公式。 对二维矩阵输入和输出变量执行二次回归函数的计算机仿真。

关键字：回归函数，参数估计，最大似然估计，贝叶斯估计，多维矩阵

Abstract. *The problem of the parameters estimation for the polynomial in the input variables regression function is formulated and solved. The input and output variables of the regression function are multidimensional-matrices. The parameters of the regression function are assumed to be random independent multidimensional matrices with Gaussian distribution and known mean value and dispersion matrices. The solution to this problem is a multidimensional-matrix system of the linear algebraic equations in multidimensional-matrix unknowns – function regression parameters. We have considered particular case of quadratic regression function, for which we have obtained formulas for parameters calculation. The computer simulation of the quadratic regression functions is performed for the two-dimensional matrix input and output variables.*

Keywords: *regression function, parameters estimations, maximum likelihood estimations, Bayesian estimations, multidimensional matrices*

1. Introduction

To date the most popular methods to estimate the parameters of the regression function are maximum likelihood method and least squares method [1, 2]. The estimations obtained by this method have good asymptotic properties and it is the justification to their application. But usage of classical methods becomes problematic in the case of a small size of the sample. In this connection Bayesian method to estimate the parameters of the regression function is more attractive. The possibility of using the samples with a small size is a significant advantage of the Bayesian approach. Interest to Bayesian inference lies in the problem of optimal (dual) control of regression objects [3, 4, 5], in econometrics [6], in other areas [7, 8]. The existing investigations into Bayesian approach relate mainly to linear in the parameters and in the input variables regression functions. There are also more general results. So, in work [5] Bayesian estimations of the parameters were obtained for the regression function represented as a scalar product of the parameter vector and the vector of the basis functions. Such representation is applicable to both linear and nonlinear regression functions in the input variables. However, such an approach is bad formalized and do not has the algorithmic generality; i. e. the mathematical expression for the vector of the basis functions is not determined and the software implementation is inapplicable for any number of variables and any degree of the polynomial.

In the present paper we investigate a multidimensional-matrix polynomial in the input variables regression function. In this case there are not the disadvantages pointed above. Such an effect is achieved by the new multidimensional-matrix representation of the polynomial regression function.

2. Problem statement

Let us consider some object with q-dimensional-matrix input variable $x=(x_j)$, $j=(j_1, j_2,..., j_q)$, p - dimensional-matrix output variable $\eta=(\eta_i)$, $i=(i_1,i_2,...,i_p)$ [9, 10], and suppose that output variable η is stochastically dependent on input variable χ so there is conditional probability density $f(\eta/x)$. We denote $y=\varphi(x)$ regression function η on χ and assume that dependence η on χ could be represented in the form $\eta=\varphi(x)+\varepsilon$, where ε is p-dimensional random matrix. Let for some values $x_1, x_2,..., x_n$ of input variable χ we obtained the values $y_{o,1}, y_{o,2},..., y_{o,n}$ of output variable η (observations, measurements) as follows:

$$y_{o,\mu}=\varphi(x_\mu)+z_\mu,\ \mu=1,...,n, \tag{1}$$

where z_μ is a realization of the random matrix ε, which we will name as errors of the measurements. We will consider the Gaussian distribution of the matrix ε with zero mean value and dispersion matrix R_ε.

Here and below we will use the following notations for indices of multidimensional matrices: $i_1,i_2,...$, are separate indices, $\bar{i}_{(p)}=(i_1,i_2,...,i_p)$ is a set of p indices (p - multiindex), $\bar{\bar{i}}_{(p,k)}=(\bar{i}_{(p),1},\bar{i}_{(p),2},...,\bar{i}_{(p),k})$ is a set of k p-multiindices.

Let the hypothetic regression function be the polynomial of m-th degree [10]:

$$\varphi(x)=\sum_{k=0}^{m}{}^{0,kq}(C_{(p,kq)}x^k)=\sum_{k=0}^{m}{}^{0,kq}(x^k C_{(kq,p)}),\ m=0,1,2,..., \tag{2}$$

where $C_{(p,kq)}$ and $C_{(kq,p)}$ are multidimensional-matrix parameters of the regression function, $C_{(p,kq)}$ is $(p+kq)$ - multidimensional matrix:

$$C_{(p,kq)} = (c_{\bar{i}_{(p)},\bar{\bar{j}}_{(q,k)}}),\ \bar{i}_{(p)} = (i_1, i_2, ..., i_p),\ \bar{\bar{j}}_{(q,k)} = (\bar{j}_{(q)1}, \bar{j}_{(q)2}, ..., \bar{j}_{(q)k}).$$

It is symmetric relative to p-multiindices $\bar{j}_{(q)1}, \bar{j}_{(q)2}, ..., \bar{j}_{(q)k}$. The matrix $C_{(kq,p)}$ is transposed matrix $C_{(p,kq)}$, i. e.

$$C_{(p,kq)} = (C_{(kq,p)})^{H_{p+kq,kq}},\ C_{(kq,p)} = (C_{(p,kq)})^{B_{p+kq,kq}},$$

where $H_{p+kq,kq}$ and $B_{p+kq,kq}$ are transpose substitutions of the types 'back' and 'onward' respectively [10]. We also denote ${}^{0,kq}(C_{(p,kq)}x^k)$ $(0,kq)$-rolled product of matrices $C_{(p,kq)}$, kq is product between k and q, $x^k = {}^{0,0}x^k$ is the (0,0) - rolled degree of the matrix x [9, 10].

In these conditions the measurement $y_{o,\mu}$ (1) has the probability density

$$f(y_{o,\mu} / x_\mu, C_{(p,0)}, C_{(p,q)}, ..., C_{(p,mq)}) =$$
$$= C_y \exp\left(-\frac{1}{2}\, {}^{0,2p}(R_\varepsilon^{-1}(y_{0,\mu} - \sum_{k=0}^{m} {}^{0,kq}(C_{(p,kq)}x_\mu^k))^2) \right),\ \mu = 1, ..., n, \quad (3)$$

where C_y is a normalizing constant, R_ε^{-1} is $(0, p)$ - inverse to R_ε matrix [9, 10].

The problem consists in finding the estimations of parameters $C_{(p,kq)}$ $(C_{(kq,p)})$ of the regression function (2) by using the given measurements $(x_1, y_{o,1})$, $(x_2, y_{o,2}), ..., (x_n, y_{o,n})$.

3. Bayesian multidimensional-matrix polynomial empirical regression

In addition to the assumptions (1)–(3) we will consider the parameter $C_{(p,kq)}$ of multidimensional-matrix polynomial regression (2) as a random matrix with Gaussian priori probability density

$$f_a(C_{(p,kq)}) = K_{(p,kq)} \exp(-\frac{1}{2}({}^{0,2(p+kq)}(R^{-1}_{a,(p,kq)}(C_{(p,kq)} - C_{a,(p,kq)})^2),\quad (4)$$

$$k = 0,1,2,...,m,\ m = 0,1,2,...,$$

where $K_{(p,kq)}$ is a normalizing constant, $C_{a,(p,kq)} = (C_{a,(kq,p)})^{H_{p+kq,kq}}$, $C_{a,(kq,p)} = (C_{a,(p,kq)})^{B_{p+kq,kq}}$ is a priori mean value $(p+kq)$ - dimensional matrix), $R_{a,(p,kq)} = (R_{a,(kq,p)})^{(H_{p+kq,kq}, H_{p+kq,kq})}$, $R_{a,(kq,p)} = (R_{a,(p,kq)})^{(B_{p+kq,kq}, B_{p+kq,kq})}$, is a priori dispersion matrix $(2(p+kq)$ - dimensional matrix), $R^{-1}_{a,(p,kq)} = (R^{-1}_{a,(kq,p)})^{(H_{p+kq,kq}, H_{p+kq,kq})}$, $R^{-1}_{a,(kq,p)} = (R^{-1}_{a,(p,kq)})^{(B_{p+kq,kq}, B_{p+kq,kq})}$, are $(0, p+kq)$ -inverse to the $R_{a,(p,kq)}$, $R_{a,(kq,p)}$ matrices respectively. We will assume that the parameters $C_{(p,0)}$, $C_{(p,q)}, \ldots,$ $C_{(p,mq)}$ are independent, i. e.

$$f_a(C_{(p,0)},..., C_{(p,mq)}) = \prod_{k=0}^{m} f_a(C_{(p,kq)}). \tag{5}$$

In these assumptions on the base of measurements $(x_1, y_{o.1}), (x_2, y_{o.2}),...,$ $(x_n, y_{o,n})$ we will find the Bayesian estimations $\hat{C}_{(p,0)}$, $\hat{C}_{(p,q)}$, …, $\hat{C}_{(p,mq)}$ of the unknown parameters $C_{(p,0)}$, $C_{(p,q)}$ …, i. e. the estimations minimizing the average risk:

$$r = E(W(C_{(p,0q)},..., C_{(p,mq)}, \hat{C}_{(p,0q)},..., \hat{C}_{(p,mq)})),$$

where $W(C_{(p,0q)},..., C_{(p,mq)}, \hat{C}_{(p,0q)},..., \hat{C}_{(p,mq)})$ is loss function, *E* is a symbol of mathematical expectation.

Theorem. Under conditions (1)–(3), (4), (5) relative to the multidimensional-matrix polynomial regression and quadratic loss function the Bayesian estimations $\hat{C}_{(p,0)}$, $\hat{C}_{(p,q)}, \ldots, \hat{C}_{(p,mq)}$ of the parameters $C_{(p,0)}$, $C_{(p,q)}, \ldots, C_{(p,mq)}$ satisfy the following system of linear multidimensional-matrix equations:

$$^{0,(p+\lambda q)}(R^{-1}_{a,(p,\lambda q)} C_{(p,\lambda q)}) + \sum_{k=0}^{m} {}^{0,(p+kq)}(V^{T_{k,\lambda}}_{k,\lambda} C_{(p,kq)}) =$$
$$= {}^{0,p}(R^{-1}_{\varepsilon} S_{yx^{\lambda}}) + {}^{0,(p+\lambda q)}(R^{-1}_{a,(p,\lambda q)} C_{a,(p,\lambda q)}),\ \lambda = 0,1,..., m, \tag{6}$$

where $S_{yx^{\lambda}}$ and $S_{x^k x^{\lambda}}$ are defined by expressions

$$S_{yx^{\lambda}} = \sum_{\mu=1}^{n} {}^{0,0}(y_{\mu} x^{\lambda}_{\mu}),\ S_{x^k x^{\lambda}} = \sum_{\mu=1}^{n} {}^{0,0}(x^k_{\mu} x^{\lambda}_{\mu}),$$

$V_{k,\lambda}$ is $(2p + kq + \lambda q)$-dimensional matrix,

$$V_{k,\lambda} = {}^{0,0}(R^{-1}_{\varepsilon} S_{x^k x^{\lambda}}),$$

R^{-1}_{ε} is (0, *p*)-inverse to the R_{ε} matrix, $V^{T_{k,\lambda}}_{k,\lambda}$ is transposed in accordance with substitution $T_{k,\lambda}$ matrix $V_{k,\lambda}$, and

$$T_{k,\lambda} = \begin{pmatrix} \bar{i}_{(p)}, \bar{\bar{v}}_{(q,\lambda)}, \bar{j}_{(p)}, \bar{\bar{t}}_{(q,k)} \\ \bar{i}_{(p)}, \bar{j}_{(p)}, \bar{\bar{t}}_{(q,k)}, \bar{\bar{v}}_{(q,\lambda)} \end{pmatrix}.$$

We do not provide a proof of the theorem.

4. Bayesian multidimensional-matrix quadratic empirical regression

Assumption $m = 2$ in the expression (2) gives us quadratic regression function:

$$y = C_{(p,0q)} + {}^{0,p}(C_{(p,1q)}x) + {}^{0,2p}(C_{(p,2q)}x^2). \quad (7)$$

In works [11, 12] one can find the algorithm of calculation of the ML-estimations of the parameters $C_{(p,0q)}$, $C_{(p,1q)}$, $C_{(p,2q)}$ for the regression function (7). Here we obtain the Bayesian estimations of these parameters.

The system of equations (6) for these parameters contains three equations:

$${}^{0,(p+0q)}(R^{-1}_{a,(p,0q)}C_{(p,0q)}) + {}^{0,(p+0q)}(V^{T_{0,0}}_{0,0}C_{(p,0q)}) + {}^{0,(p+1q)}(V^{T_{1,0}}_{1,0}C_{(p,1q)}) + {}^{0,(p+2q)}(V^{T_{2,0}}_{2,0}C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^0}) + {}^{0,(p+0q)}(R^{-1}_{a,(p,0q)}C_{a,(p,0q)}),$$

$${}^{0,(p+1q)}(R^{-1}_{a,(p,1q)}C_{(p,1q)}) + {}^{0,(p+0q)}(V^{T_{0,1}}_{0,1}C_{(p,0q)}) + {}^{0,(p+1q)}(V^{T_{1,1}}_{1,1}C_{(p,1q)}) + {}^{0,(p+2q)}(V^{T_{2,1}}_{2,1}C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^1}) + {}^{0,(p+1q)}(R^{-1}_{a,(p,1q)}C_{a,(p,1q)}),$$

$${}^{0,(p+2q)}(R^{-1}_{a,(p,2q)}C_{(p,2q)}) + {}^{0,(p+0q)}(V^{T_{0,2}}_{0,2}C_{(p,0q)}) + {}^{0,(p+1q)}(V^{T_{1,2}}_{1,2}C_{(p,1q)}) + {}^{0,(p+2q)}(V^{T_{2,2}}_{2,2}C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^2}) + {}^{0,(p+2q)}(R^{-1}_{a,(p,2q)}C_{a,(p,2q)}).$$

Collecting similar terms we obtain the following system of equations:

$${}^{0,p}((R^{-1}_{a,(p,0q)} + V^{T_{0,0}}_{0,0})C_{(p,0q)}) + {}^{0,(p+q)}(V^{T_{1,0}}_{1,0}C_{(p,1q)}) + {}^{0,(p+2q)}(V^{T_{2,0}}_{2,0}C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^0}) + {}^{0,p}(R^{-1}_{a,(p,0q)}C_{a,(p,0q)}),$$

$${}^{0,p}(V^{T_{0,1}}_{0,1}C_{(p,0q)}) + {}^{0,(p+q)}((R^{-1}_{a,(p,1q)} + V^{T_{1,1}}_{1,1})C_{(p,1q)}) + {}^{0,(p+2q)}(V^{T_{2,1}}_{2,1}C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^1}) + {}^{0,(p+q)}(R^{-1}_{a,(p,1q)}C_{a,(p,1q)}),$$

$${}^{0,p}(V^{T_{0,2}}_{0,2}C_{(p,0q)}) + {}^{0,(p+q)}(V^{T_{1,2}}_{1,2}C_{(p,1q)}) + {}^{0,(p+2q)}((R^{-1}_{a,(p,2q)} + V^{T_{2,2}}_{2,2})C_{(p,2q)}) =$$

$$= {}^{0,p}(R^{-1}_{\varepsilon}S_{yx^2}) + {}^{0,(p+2q)}(R^{-1}_{a,(p,2q)}C_{a,(p,2q)}).$$

With notations

$$R_{(p,0q)}=(R_{a,(p,0q)}^{-1}+V_{0,0}^{T_{0,0}})\,,$$

$$R_{(p,1q)}=(R_{a,(p,1q)}^{-1}+V_{1,1}^{T_{1,1}})\,,$$

$$R_{(p,2q)}=(R_{a,(p,2q)}^{-1}+V_{2,2}^{T_{2,2}})\,,$$

$$B_{(p)}={}^{0,p}(R_{\varepsilon}^{-1}\,S_{yx^0})+{}^{0,p}(R_{a,(p,0q)}^{-1}C_{a,(p,0q)})\,,$$

$$B_{(p+q)}={}^{0,p}(R_{\varepsilon}^{-1}\,S_{yx^1})+{}^{0,(p+q)}(R_{a,(p,1q)}^{-1}C_{a,(p,1q)})\,,$$

$$B_{(p+2q)}={}^{0,p}(R_{\varepsilon}^{-1}\,S_{yx^2})+{}^{0,(p+2q)}(R_{a,(p,2q)}^{-1}C_{a,(p,2q)})\,,$$

,

we rewrite this system in the form:

$$\begin{cases} {}^{0,p}(R_{(p,0q)}C_{(p,0q)})+{}^{0,(p+q)}(V_{1,0}^{T_{1,0}}C_{(p,1q)})+{}^{0,(p+2q)}(V_{2,0}^{T_{2,0}}C_{(p,2q)})=B_{(p)}, \\ {}^{0,p}(V_{0,1}^{T_{0,1}}C_{(p,0q)})+{}^{0,(p+q)}(R_{(p,1q)}C_{(p,1q)})+{}^{0,(p+2q)}(V_{2,1}^{T_{2,1}}C_{(p,2q)})=B_{(p+q)}, \\ {}^{0,p}(V_{0,2}^{T_{0,2}}C_{(p,0q)})+{}^{0,(p+q)}(V_{1,2}^{T_{1,2}}C_{(p,1q)})+{}^{0,(p+2q)}(R_{(p,2q)}C_{(p,2q)})=B_{(p+2q)}. \end{cases}$$

This system of equations can be solved by the elimination method.

5. Conclusion

In conclusion, we outline the main results of this work and note their particularities.

1. The problem of the building of the Bayesian multidimensional-matrix polynomial regression was formulated and solved. This regression has the following particularities compared with existing regressions: 1) the more general multidimensional-matrix polynomial regression function, when input and output variables are the multidimensional matrices, is considered; 2) a new untraditional multidimensional-matrix form of the representation of the regression function in the manner of multidimensional-matrix polynomial is used. Besides, the priory distributions of the multidimensional-matrix parameters of the regression function are supposed as Gaussian. The general solution of this problem is the system of the linear multidimensional-matrix equations relative the multidimensional-matrix parameters of the regression function.

2. On the base of the general solution the algorithm of the parameters calculation of the Bayesian multidimensional-matrix quadratic empirical regression functions was obtained.

3. The computer simulation of the quadratic Bayesian empirical regressions function with two-dimensional input and output variables was performed. The simulation confirmed the correctness of the theoretical results and illustrated the important benefits of the Bayesian approach to have the algorithmic generality and to obtain the estimations on the small number of the measurements.

References

1. S.M. Ermakov, A.A. Zhiglyavskii, Mathematical theory of optimal experiment. Moscow, Publishing house "Science", 1987. 320 p. In Russian.

2. N.P. Klepikov, S.N. Sokolov, Analysis and planning of experiments by the method of maximum likelihood. Berlin, Akademie-Verlag, 1961. 127 p.

3. A.A. Feldbaum, The bases of the theory of optimal automatic systems. Moscow, Publishing house "Science", 1966. 552 p. In Russian.

4. V.S. Mukha, E.V. Sergeev, Dual control of regression objects. Proceedings of LETI. Issue 202, 1976. Pp. 58–65. In Russian.

5. V.S. Mukha, E.V. Sergeev, Bayesian estimations of the parameters of the regression objects. Proceedings of LETI. Issue 149, 1974. Pp. 26–29. In Russian.

6. A. Zellner, An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, Inc. 1971. 448 p.

7. Ludwig Fahrmeir, Thomas Kneib, Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data. Oxford University Press, Oxford, 2011. 544 p.

8. Jon Wakefield, Bayesian and Frequentist Regression Methods. Springer, 2013. 709 p.

9. N.P. Socolov, Introduction to the theory of multidimensional matrices. Kiev, Naukova dumka, 1972. 176 p. In Russian.

10. V.S. Mukha, Analysis of multidimensional data. Minsk, Technoprint, 2004. 368 p. In Russian.

11. V.S. Mukha, Multidimensional-matrix polynomial regression analysis. Parameters estimations. Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series. 2007, No. 1, pp. 45–51. In Russian.

12. V.S. Mukha, The best polynomial multidimensional-matrix regression. Cybernetics and Systems Analysis. Vol. 43, Issue 3, 2007. Pp 427–432.