Conditional Multidimensional Parameter Identification with Asymmetric Correlated Losses of Estimation Errors

Piotr Kulczycki and Malgorzata Charytanowicz

Polish Academy of Sciences, Systems Research Institute
Centre of Information Technology for Data Analysis Methods
Piotr.Kulczycki@ibspan.waw.pl

Abstract. This paper is dedicated to the problem of the estimation of a vector of parameters, as losses resulting from their under- and overestimation are asymmetric and mutually correlated. The issue is considered from an additional conditional aspect, where particular coordinates of conditioning variables may be continuous, binary, discrete or categorized (ordered and unordered). The final result is an algorithm for calculating the value of an estimator, optimal in sense of expectation of losses using a multidimensional asymmetric quadratic function, for practically any distributions of describing and conditioning variables.

Keywords: parameters' vector identification, conditional factors, Bayes approach, asymmetric loss function, distribution free method, numerical algorithm.

1 Introduction

The proper identification (estimation) of parameters values, used in a model describing the reality under consideration, is always of fundamental significance in modern problems of science and practice. The need to consider implications of estimations errors different for under- and overestimations, leads directly to the concept of asymmetrical form of a loss function [Berger, 1980]. The significance of this problem has been investigated for simple cases of a single parameter [Zellner, 1985; McCullough, 2000]. It is also worth noting the results concerning the estimation of a single parameter with asymmetrical loss function, described in the paper [Kulczycki and Charytanowicz, 2013] in the conditional version, i.e. where the quantity under research is significantly dependent on conditional factors. If the actual value of factors of this type is available metrologically, their inclusion can make the model used considerably more precise. In this paper that research is generalized for the multidimensional case, where one identifies a few independent parameters, treated as a vector, and the losses resulting from the over- and underestimation may be asymmetrical and correlated. The concept presented here is based on the Bayes approach, which allows minimization of expected value of losses arising from estimation errors. For defining probability characteristics, the nonparametric methodology of statistical kernel

L.C. Kiong et al. (Eds.): ICONIP 2014, Part II, LNCS 8835, pp. 287–294, 2014.

estimators was used, which freed the investigated procedure from forms of distributions characterizing both the identified parameters and conditioning quantities.

2 Statistical Kernel Estimators

Let the n-dimensional random variable X be given, with a distribution characterized by the density f. Its kernel estimator $\hat{f}: \mathbb{R}^n \to [0,\infty)$, calculated using experimentally obtained values for the m-element random sample

$$x_1, x_2, \ldots, x_m, \tag{1}$$

in its basic form is defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{2}$$

where $m \in \mathbb{N} \setminus \{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K: \mathbb{R}^n \to [0,\infty)$ of unit integral $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. The method presented here uses the one-dimensional Cauchy kernel, in the n-dimensional case generalized to the product kernel. For fixing the smoothing parameter value, the plug-in method is recommended, with the modification of this parameter. Details are found in the books [Kulczycki, 2005; Silverman, 1986; Wand and Jones, 1994].

The above concept will now be generalized for the conditional case. Here, besides the basic (termed the describing) n_Y-dimensional random variable Y, let also be given the n_W-dimensional random variable W, called hereinafter the conditioning random variable. Their composition $X = \begin{bmatrix} Y \\ W \end{bmatrix}$ is a random variable of the dimension $n_Y + n_W$. Assume that distributions of the variables X and, in consequence, W have densities, denoted below as $f_X : \mathbb{R}^{n_Y + n_W} \to [0,\infty)$ and $f_W : \mathbb{R}^{n_W} \to [0,\infty)$, respectively. Let also be given the so-called conditioning value, that is the fixed value of conditioning random variable $w^* \in \mathbb{R}^{n_W}$ such that $f_W(w^*) > 0$, and also the random sample

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \ldots, \begin{bmatrix} y_m \\ w_m \end{bmatrix}, \tag{3}$$

obtained from the variable X. The particular elements of this sample are interpreted as the values y_i taken in measurements from the random variable Y, when the conditioning variable W assumes the respective values w_i. The kernel estimator of conditional density of the random variable Y distribution for the conditioning value w^*, i.e. $\hat{f}_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0,\infty)$, can be given by the following form helpful in practice:

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix}\right) = \tag{4}$$

$$= \frac{1}{h_1 h_2 \ldots h_{n_Y} \sum_{i=1}^{m} d_i} \sum_{i=1}^{m} d_i K_1\left(\frac{y_1 - y_{i,1}}{h_1}\right) K_2\left(\frac{y_2 - y_{i,2}}{h_2}\right) \ldots K_{n_Y}\left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right),$$

where h_1, $h_2, \ldots, h_{n_Y+n_W}$ represent smoothing parameters for particular coordinates of the random variable X, while the coordinates of the vectors w^*, y_i and w_i are denoted as

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W}^* \end{bmatrix} \quad \text{and} \quad y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_Y} \end{bmatrix}, \quad w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{for} \quad i = 1, 2, \ldots, m, \tag{5}$$

whereas the so-called conditioning parameters d_i for $i = 1, 2, \ldots, m$ can be defined by

$$d_i = K_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) K_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots K_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right). \tag{6}$$

The value of the parameter d_i characterizes the "distance" of the given conditioning value w^* from w_i – that of the conditioning variable for which the i-th element of the random sample was obtained. Then estimator (4) can be interpreted as the linear combination of kernels mapped to particular elements of a random sample obtained for the variable Y, when the coefficients of this combination characterize how representative these elements are for the given value w^*.

3 An Algorithm

Consider the parameters, whose values are to be estimated, denoted in the form of the vector $y \in \mathbb{R}^{n_Y}$. It will be treated as the value of the n_Y-dimensional random variable Y. Let also the n_W-dimensional conditional random variable W be given. The availability is assumed of the metrologically achieved measurements of the parameters' vector y, i.e. y_1, $y_2, \ldots, y_m \in \mathbb{R}^{n_Y}$, obtained for the values w_1, $w_2, \ldots, w_m \in \mathbb{R}^{n_W}$ of the conditional variable, respectively. Finally, let $w^* \in \mathbb{R}^{n_W}$ denote any fixed conditioning value. The goal is to calculate the estimator of this parameter's vector, optimal in the sense of minimum expected value of losses arising from errors of estimation, for conditioning value w^*. In order to solve such a task, the Bayes decision rule will be used. Because of clarity of presentation, a two-dimensional case ($n_Y = 2$) will be considered here. The idea itself may be transposed for larger dimensions, although at a natural – in such a situation – cost of increasing complexity.

Let therefore the estimated parameters be treated as the two-dimensional vector $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, as their estimators $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$. The loss function $l : \mathrm{R}^2 \times \mathrm{R}^2 \to \mathrm{R}$, which in accordance with the decision theory principles [Berger, 1980] defines losses occurring when the value $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$ has been taken, while in reality the hypothetical state was $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, is assumed in a quadratic and asymmetrical form:

$$l\left(\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{cases} a_l(\hat{y}_1 - y_1)^2 + a_{ld}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \qquad \text{if} \quad \hat{y}_1 - y_1 \le 0 \text{ and } \hat{y}_2 - y_2 \le 0 \\ a_r(\hat{y}_1 - y_1)^2 + a_{rd}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \qquad \text{if} \quad \hat{y}_1 - y_1 \ge 0 \text{ and } \hat{y}_2 - y_2 \le 0 \\ a_l(\hat{y}_1 - y_1)^2 + a_{lu}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \qquad \text{if} \quad \hat{y}_1 - y_1 \le 0 \text{ and } \hat{y}_2 - y_2 \ge 0 \\ a_r(\hat{y}_1 - y_1)^2 + a_{ru}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \qquad \text{if} \quad \hat{y}_1 - y_1 \ge 0 \text{ and } \hat{y}_2 - y_2 \ge 0 \end{cases} . \tag{7}$$

where $a_l, a_r, a_u, a_d > 0$, $a_{ld}, a_{ru} \ge 0$ and $a_{lu}, a_{rd} \le 0$. The coefficients a_{ld}, a_{ru}, a_{lu}, a_{rd} represent the complementary correlation of estimation errors for both parameters.

Assume conditional independence [Dawid, 1979] of the estimated parameters. Then the density $f_{Y|W=w^*}$ representing their uncertainty may be shown as the product of the one-dimensional densities $f_{Y_1|W=w^*} : \mathrm{R} \to [0,\infty)$ and $f_{Y_2|W=w^*} : \mathrm{R} \to [0,\infty)$ corresponding to particular composites, i.e.

$$f_{Y|W=w^*}(y_1, y_2) = f_{Y_1|W=w^*}(y_1) f_{Y_2|W=w^*}(y_2) . \tag{8}$$

Let also the functions $f_{Y_1|W=w^*}$ and $f_{Y_2|W=w^*}$ be continuous and such that $\int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 < \infty$ as well as $\int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 < \infty$.

Detailed analysis [Kulczycki and Charytanowicz, 2014] shows that the criterion minimizing the expectation value of losses takes on the form of equations unsolvable in practice. Although if estimation of the densities presented above is reached using the kernel estimators, then one can design an effective numerical algorithm to this end. Thus, with any fixed $i = 1, 2, \ldots, m$, one can define the functions $U_{1,i} : \mathrm{R} \to \mathrm{R}$, $U_{2,i} : \mathrm{R} \to \mathrm{R}$, $V_{1,i} : \mathrm{R} \to \mathrm{R}$ and $V_{2,i} : \mathrm{R} \to \mathrm{R}$, given as

$$U_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} K\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathrm{d}y_1 \tag{9}$$

$$U_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} K\left(\frac{y_2 - y_{i,2}}{h_2}\right) \mathrm{d}y_2 \tag{10}$$

$$V_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} y_1 K\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathrm{d}y_1 \tag{11}$$

$$V_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} y_2 K\left(\frac{y_2 - y_{i,2}}{h_2}\right) \mathrm{d}y_2 . \tag{12}$$

Norm also the conditioning parameters d_i by introducing the positive values

$$d_i^* = \frac{d_i}{\sum_{i=1}^m d_i} \qquad \text{for} \quad i = 1, 2, \ldots, m ; \tag{13}$$

note that $\sum_{i=1}^m d_i^* = 1$. Then criterion for the vector $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$ ensuring the minimum of expectation value of losses for loss function (7) takes the form of the equations

$$\begin{aligned}
&\sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1)\left[(a_{ru} - a_{rd} - a_{lu} + a_{ld})\sum_{i=1}^m d_i^*\big(\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)\big) + (a_{rd} - a_{ld})(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2})\right] + \\
&\quad + 2a_l(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1}) + 2(a_r - a_l)\sum_{i=1}^m d_i^*\big(\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)\big) + a_{ld}(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2}) + \\
&\quad + (a_{lu} - a_{ld})\sum_{i=1}^m d_i^*\big(\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)\big) = 0
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2)\left[(a_{ru} - a_{rd} - a_{lu} + a_{ld})\sum_{i=1}^m d_i^*\big(\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)\big) + (a_{lu} - a_{ld})(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1})\right] + \\
&\quad + 2a_d(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2}) + 2(a_u - a_d)\sum_{i=1}^m d_i^*\big(\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)\big) + a_{ld}(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1}) + \\
&\quad + (a_{rd} - a_{ld})\sum_{i=1}^m d_i^*\big(\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)\big) = 0 .
\end{aligned} \tag{15}$$

If one denotes the left sides of the above equations as $L_1(\hat{y}_1, \hat{y}_2)$ and $L_2(\hat{y}_1, \hat{y}_2)$, their partial derivatives are given by

$$\frac{\partial L_1(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_1}=\sum_{i=1}^{m} d_i^* \frac{1}{h_1 s_{i,1}} K\left(\frac{\hat{y}_1-y_{i,1}}{h_1 s_{i,1}}\right)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^*\big(\hat{y}_2 U_{2,i}(\hat{y}_2)-V_{2,i}(\hat{y}_2)\big)+\right.$$

$$\left.+(a_{rd}-a_{ld})(\hat{y}_2-\sum_{i=1}^{m} d_i^* y_{i,2})\right]+2(a_r-a_l)\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)+2a_l \tag{16}$$

$$\frac{\partial L_1(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2}=\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+(a_{rd}-a_{ld})\right]+$$

$$+(a_{lu}-a_{ld})\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+a_{ld} \tag{17}$$

$$\frac{\partial L_2(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_1}=\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)+(a_{lu}-a_{ld})\right]+$$

$$+(a_{rd}-a_{ld})\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)+a_{ld} \tag{18}$$

$$\frac{\partial L_2(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2}=\sum_{i=1}^{m} d_i^* \frac{1}{h_2 s_{i,2}} K\left(\frac{\hat{y}_2-y_{i,2}}{h_2 s_{i,2}}\right)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^*\big(\hat{y}_1 U_{1,i}(\hat{y}_1)-V_{1,i}(\hat{y}_1)\big)+\right.$$

$$\left.+(a_{lu}-a_{ld})(\hat{y}_1-\sum_{i=1}^{m} d_i^* y_{i,1})\right]+2(a_u-a_d)\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+2a_d\,. \tag{19}$$

Then the solution of equations (14)-(15) can be calculated through Newton's multidimensional algorithm [Stoer and Bulirsch, 2002] as the limit of the two-dimensional sequence $\left\{\begin{matrix}\hat{y}_{j,1}\\ \hat{y}_{j,2}\end{matrix}\right\}_{j=0}^{\infty}$ defined by formulas

$$\hat{y}_{0,1}=\frac{\sum_{i=1}^{m} d_i\, y_{i,1}}{\sum_{i=1}^{m} d_i} \qquad \text{and} \qquad \hat{y}_{0,2}=\frac{\sum_{i=1}^{m} d_i\, y_{i,2}}{\sum_{i=1}^{m} d_i} \tag{20}$$

$$\begin{bmatrix}\hat{y}_{j+1,1}\\ \hat{y}_{j+1,2}\end{bmatrix}=\begin{bmatrix}\hat{y}_{j,1}\\ \hat{y}_{j,2}\end{bmatrix}-\begin{bmatrix}\dfrac{\partial L_1(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_1} & \dfrac{\partial L_1(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_2}\\ \dfrac{\partial L_2(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_1} & \dfrac{\partial L_2(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_2}\end{bmatrix}^{-1}\begin{bmatrix}L_1(\hat{y}_{j,1},\hat{y}_{j,2})\\ L_2(\hat{y}_{j,1},\hat{y}_{j,2})\end{bmatrix}$$
$$\text{for} \quad j=0,1,\ldots, \tag{21}$$

while the quantities in the above dependencies are given by equations (14)-(19), whereas a stop condition takes the form of the conjunction of the inequalities

$$|\,\hat{y}_{j,1}-\hat{y}_{j-1,1}\,|\le 0{,}01\,\hat{\sigma}_1 \qquad \text{and} \qquad |\,\hat{y}_{j,2}-\hat{y}_{j-1,2}\,|\le 0{,}01\,\hat{\sigma}_2\,, \tag{22}$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ denote the estimators of standard deviations for particular coordinates of the vector Y.

4 Final Comments and Summary

This paper presents the algorithm for calculating the conditional estimator of the vector of independent parameters, where losses resulting from under- and overestimation are asymmetrical and mutually correlated. The conditional approach allows in practice for refinement of the model by including the current value of the conditioning factors. Use of the Bayes approach ensures a minimum expected value of losses, a statistical kernel estimators methodology frees the investigated procedure from forms of distributions of the describing and conditioning factors.

The correct performance of the algorithm has been proved in many numerical tests with illustrative generated data, and also by simulations, as well as by applying them to practical problems from control engineering, biomedicine and marketing. Above all, the general rule was translations of the examined estimator values in directions associated with smaller losses resulting from estimation errors, defined by loss function (7). Specifically, an increase in the value of the parameter a_l with respect to a_r, and so a growth in the value of this function for positive estimation errors of the first parameter (its overestimation), implied an increase in the value of the obtained estimator for this parameter. In consequence it reduces the probability of an overestimation. The opposite occurs when the parameter a_r value is increased with respect to a_l: the value of the obtained estimator decreases, which lowers the probability of underestimation. The more the ratio a_l/a_r differed from 1, the more intensive are the above effects. Analogous dependences appeared for the parameters a_d and a_u when estimating the second parameter. Subsequently the increase in the parameter a_{ld} value resulted in the simultaneous growth in the two parameters' values, reducing in both cases the probability of overestimation. Converse effects implied changes in the parameter a_{ru}. And finally, an increase in the absolute value of the parameter a_{rd} reduces the probability of underestimation of the first parameter as well as overestimation of the second, through a decrease in the value of the obtained estimator for the first, and an increase for the second. The opposite applies to the parameter a_{lu}.

The conditional approach implied the appropriate correction to the estimator value according to the nature of the correlation between describing variables and conditioning factors. If a parameter was positively correlated to such factor, an increase/decrease in the condition value resulted in an increase/decrease in the estimator value for that parameter. The opposite occurred for a negative correlation. Such relation may be more complex, according to any potential form of the dependence of the conditional densities $f_{Y_1|W=w^*}$ and $f_{Y_2|W=w^*}$ on conditioning values w^*.

An acceptable quality of results was obtained from sample sizes of just 50-100 when the conditioning value was positioned close to the main modal value of a conditional variable, and 100-200 at distance of standard deviation. Taking into account the complex multidimensional character of the task, it does not seem to be an excessive requirement in practice. Thanks to the averaging properties of kernel estimators, the algorithm proved to be robust to a small number or even lack of data from the neighborhood of a conditioning value.

The procedure presented in this paper has been given in its basic form. A clear interpretation means it is possible to make individual modifications. Above all this allows the inclusion of conditional factors other than continuous (real). Similarly to the kernel estimation definition formulated above for continuous random variables, one can construct kernel estimators for binary, discrete and categorized (including ordered) variables, as well as any of their compositions, especially with continuous variables – for details see broad and varied literature in this subject. The above can be particularly useful for the modern data analysis tasks, which more and more often take advantage of the many different configurations for particular types of attributes.

A broad description of the methodology introduced in this paper is presented in the article [Kulczycki and Charytanowicz, 2014] together with results of detailed verifications. The algorithm itself is given there in its ready-to-use form and can be applied directly without deep subject knowledge or laborious research.

Acknowledgments. Our heartfelt thanks go to our colleague Dr. Aleksander Mazgaj, with whom we commenced the research presented here. With his consent, this tex t also contains results of joint research.

References

1. Berger, J.O.: Statistical Decision Theory. Springer, New York (1980)
2. Dawid, A.P.: Conditional Independence in Statistical Theory. Journal of the Royal Statistical Society, Series B 41, 1–31 (1979)
3. Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)
4. Kulczycki, P., Charytanowicz, M.: Conditional Parameter Identification with Different Losses of Under- and Overestimation. Applied Mathematical Modelling 37, 2166–2177 (2013)
5. Kulczycki, P., Charytanowicz, M.: An Algorithm for Conditional Multidimensional Parameter Identification with Asymmetric and Correlated Looses of Under- and Overestimations (in press, 2014)
6. Lehmann, E.L.: Theory of Point Estimation. Wiley, New York (1983)
7. McCullough, B.D.: Optimal Prediction with a General Loss Function. Journal of Combinatorics, Information & System Sciences 25, ss.207–ss.221 (2000)
8. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
9. Stoer, J., Bulirsch, R.: Introduction to Numerical Analysis. Springer, New York (2002)
10. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1995)
11. Zellner, A.: Bayesian Estimation and Prediction Using Asymmetric Loss Function. Journal of the American Statistical Association 81, 446–451 (1985, 1986)