Journal of Statistical Computation and Simulation, 2016
Vol. 86, No. 5, 1032–1055, http://dx.doi.org/10.1080/00949655.2015.1049606

An algorithm for conditional multidimensional parameter identification with asymmetric and correlated losses of under- and overestimations

Piotr Kulczycki[a,b*] and Malgorzata Charytanowicz[a,c]

[a] *Polish Academy of Sciences, Systems Research Institute, Centre of Information Technology for Data Analysis Methods, Warsaw, Poland;* [b] *Faculty of Physics and Applied Computer Science, Division for Information Technology and Biometrics, AGH University of Science and Technology, Kraków, Poland;* [c] *Institute of Mathematics and Computer Science, Catholic University of Lublin, Lublin, Poland*

(*Received 23 November 2014; accepted 5 May 2015*)

Many of today's specialized applicational tasks are obliged to consider the influence of inevitable errors in the identification of parameters appearing in a model. Favourable results can also be achieved through measuring, and then accounting for definite (e.g. current) values of factors which show a significant reaction to the values of those parameters. This paper is dedicated to the problem of the estimation of a vector of parameters, where losses resulting from their under- and overestimation are asymmetric and mutually correlated. The issue is considered from a supplementary conditional aspect, where particular coordinates of conditioning variables may be continuous, discrete, multivalued (in particular binary) or categorized (ordered and unordered). The final result is a ready-to-use algorithm for calculating the value of an estimator, optimal in the sense of minimum expectation of losses using a multidimensional asymmetric quadratic function, for practically any distributions of describing and conditioning variables.

Keywords: identification of vector of parameters; conditioning factors; Bayes approach; asymmetrical loss function; nonparametric estimation; 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.[1] In contemporary complex tasks, however, the individual parameters no longer describe definite physical quantities; they rather represent an entire range of complex phenomena, simplified in the framework of a model to one parameter, existing only formally. In this situation, their identification cannot be taken in the classic sense [2] to be the calculation of an estimators' values as close as possible to the imagined 'true' – albeit unknown – parameters' values (since they do not exist), but rather by allowing for the influence of their particular values on the behaviour of a considered environment. In particular, in many modern applicational tasks, the underestimation of parameters' values may imply different results from overestimation, both in quality and quantity. The necessity for their inclusion becomes ever more attractive, with the

*Corresponding author. Emails: kulczycki@ibspan.waw.pl; kulczycki@agh.edu.pl

complexity and sophistication of contemporary models growing, as well as the demands made by the task for which they are constructed.[3]

The need to consider different implications of under- and overestimations leads directly to the concept of asymmetrical form of a loss function.[4] The significance of this problem has been noted and investigated for simple cases of a single parameter (see the classic paper [5]). An interesting comparison of basic types of loss function used then – asymmetrical linear, asymmetrical quadratic and linear-exponential LINEX (approximately linear on one side of zero and exponential on the other one) – is presented in the paper [6]. It is also worth noting the results concerning the estimation of a single parameter with asymmetrical polynomial loss function,[7] investigated in the paper [8] additionally in the conditional version, i.e. where the quantity under research is significantly dependent on conditioning factors. In engineering practice, such a factor may often be the current temperature. 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 separated 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 estimators was used, which freed the investigated procedure from forms of distributions characterizing both the identified parameters and conditioning quantities.

Finally, to summarize, the goal of this paper is the provision of an algorithm for calculating the vector of independent parameters' values, optimal in the sense of minimum expectation value of losses, when these losses are different for under- and overestimation and in addition correlated for particular parameters. The procedure is worked out for the conditional approach, which enables the result to be made more precise by a fixed (e.g. valid) value of a conditioning factor or factors. Both estimated parameter and conditioning factors can have any distribution. The algorithm will be given in its ready-to-use form, i.e. together with quoted bibliography it can be applied directly without detailed knowledge of theoretical aspects, laborious research or analytical calculations.

Thus, Section 2 outlines mathematical preliminaries: statistical kernel estimators and Bayes estimation. The algorithm worked out here is described in Sections 3 and 4. Section 5 presents the results of its numerical tests for illustrative, artificially generated data. Additional comments are provided in Section 6, while Section 7 offers as examples two applicational tasks, in which the investigated procedure can be found. The last section constitutes a summary of the designed method. Mathematical aspects are considered in the two appendixes closing the article.

The preliminary version of this paper was shortly presented as the publications.[9,10]

2. Mathematical preliminaries

2.1. *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 $m \in \mathbb{N}\backslash\{0\}$, 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 choice of form of the kernel K and the calculation of the smoothing parameter h are made most often with the criterion of the mean integrated square error.

Thus, the choice of the kernel form has – from a statistical point of view – no practical meaning and thanks to this, it becomes possible to take primarily into account properties of the estimator obtained or calculational aspects, advantageous from the point of view of the applicational problem under investigation; for broader discussion, see the books [11,Section 3.1.3;12,Sections 2.7 and 4.5]. In practice, for the one-dimensional case (i.e. when $n = 1$), the function K is assumed most often to be the density of a common probability distribution. In the multidimensional case, two natural generalizations of the above concept are used: radial and product kernels. However, the former is somewhat more effective, although from an applicational point of view, the difference is immaterial and the product kernel – significantly more convenient for analysis – is often favoured in practical problems. The n-dimensional product kernel K can be expressed as

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = K_1(x_1)\, K_2(x_2) \ldots K_n(x_n), \tag{3}$$

where K_i denotes the previously mentioned one-dimensional kernels, while the expression h^n appearing in the basic formula (2) should be replaced by the product of the smoothing parameters for particular coordinates $h_1 \cdot h_2 \cdot \ldots \cdot h_n$.

The fixing of the smoothing parameter h has significant meaning for quality of estimation. Fortunately many suitable procedures for calculating the value of the parameter h on the basis of random sample (1) have been worked out. For broader discussion of this task see the books [11–13]. In particular, for the one-dimensional case, the simple and effective plug-in method [11,Section 3.1.5;12,Section 3.6.1] is especially recommended. Of course this method can also be applied in the n-dimensional case when product kernel (3) is used, sequentially n times for each coordinate.

Practical applications may also use additional procedures generally improving the quality of estimator (2). For the method presented in this paper, the modification of the smoothing parameter [11,Section 3.1.6;13,Section 5.3.1] can be recommended.

The above concept will now be generalized for the conditional case. Here, besides the basic (sometimes 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, i.e. the fixed value of conditioning random variable $w^* \in \mathbb{R}^{n_W}$, such that

$$f_W(w^*) > 0. \tag{4}$$

Then the function $f_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0, \infty)$ given by

$$f_{Y|W=w^*}(y) = \frac{f_X(y, w^*)}{f_W(w^*)} \quad \text{for every } y \in \mathbb{R}^{n_Y} \tag{5}$$

constitutes a conditional density of probability distribution of the random variable Y for the conditioning value w^*. The conditional density $f_{Y|W=w^*}$ can so be treated as a 'classic' density, whose form has been made more accurate in practical applications with w^* – a concrete value taken by the conditioning variable W in a given situation.

Let therefore the random sample

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \dots, \begin{bmatrix} y_m \\ w_m \end{bmatrix}, \tag{6}$$

obtained from the variable $X = \begin{bmatrix} Y \\ W \end{bmatrix}$, be given. 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. On the basis of sample (6), one can calculate $\hat{f}_X$, i.e. the kernel estimator of density of the random variable X probability distribution, while the sample

$$w_1, w_2, \dots, w_m \tag{7}$$

gives $\hat{f}_W$ – the kernel density estimator for the conditioning variable W. The kernel estimator of conditional density of the random variable Y distribution for the conditioning value w^* is defined then – in natural consequence of formula (5) – as the function $\hat{f}_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0, \infty)$ given by

$$\hat{f}_{Y|W=w^*}(y) = \frac{\hat{f}_X(y, w^*)}{\hat{f}_W(w^*)}. \tag{8}$$

If for the estimator $\hat{f}_W$ one uses a kernel with positive values, then the inequality $\hat{f}_W(w^*) > 0$ implied by condition (4) is fulfilled for any $w^* \in \mathbb{R}^{n_W}$.

In the case when for the estimators $\hat{f}_X$ and $\hat{f}_W$ the product kernel (3) is used, applying in pairs the same positive kernels to the estimator $\hat{f}_X$ for coordinates which correspond to the vector W and to the estimator $\hat{f}_W$, then the expression for the kernel estimator of conditional density becomes particularly helpful for practical applications. Formula (8) can then be specified to the form

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

$$= \frac{\begin{array}{c}(1/h_1 h_2 \dots h_{n_Y}) \sum_{i=1}^m K_1((y_1 - y_{i,1})/h_1) K_2((y_2 - y_{i,2})/h_2) \dots K_{n_Y}((y_{n_Y} - y_{i,n_Y})/h_{n_Y}) \\ \times K_{n_Y+1}((w_1^* - w_{i,1})/h_{n_Y+1}) K_{n_Y+2}((w_2^* - w_{i,2})/h_{n_Y+2}) \dots K_{n_Y+n_W}((w_{n_W}^* - w_{i,n_W})/h_{n_Y+n_W})\end{array}}{\sum_{i=1}^m K_{n_Y+1}((w_1^* - w_{i,1})/h_{n_Y+1}) K_{n_Y+2}((w_2^* - w_{i,2})/h_{n_Y+2}) \dots K_{n_Y+n_W}((w_{n_W}^* - w_{i,n_W})/h_{n_Y+n_W})}, \tag{9}$$

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

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W}^* \end{bmatrix} \quad \text{and} \quad y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_Y} \end{bmatrix}, \quad w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{for } i = 1, 2, \dots, m. \tag{10}$$

Define the so-called conditioning parameters d_i for $i = 1, 2, \dots, m$ by the following formula:

$$d_i = K_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) K_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \dots K_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right). \tag{11}$$

Thanks to the assumption of positive values for the kernels $K_{n_Y+1}, K_{n_Y+2}, \ldots, K_{n_Y+n_W}$, these parameters are also positive. So the kernel estimator of conditional density (9) can be finally presented in the form

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix}\right)$$
$$= \frac{1}{h_1 h_2 \ldots h_{n_Y} \sum_{i=1}^{m} d_i} \sum_{i=1}^{m} d_i K_1\left(\frac{y_1 - y_{i,1}}{h_1}\right) K_2\left(\frac{y_2 - y_{i,2}}{h_2}\right) \ldots K_{n_Y}\left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right). \tag{12}$$

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 ith element of the random sample was obtained. Then estimator (12) 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^*.

More details concerning kernel estimators can be found in the classic monographs.[11–14]

2.2. *Bayes estimation*

The idea of Bayes estimation [2] as applied here can be derived illustratively from the Bayes decision rule common to decision theory,[4] treating the possible values of estimated parameters as states of nature, and the obtained values of estimators as a decision.

Assume therefore a nonempty set of all possible states of nature Z and f – the density of distribution of a probability measure defined on Z. Let there be given also the nonempty set of possible decisions D, as well as the loss function $l: D \times Z \to \mathbb{R}$, while its values $l(d, z)$ can be interpreted as losses occurring in a hypothetical case, when the state of nature is z and the decision d was taken. If for every $d \in D$ the integral $\int_N l(d, z) f(z) \mathrm{d}z$ exists, then the Bayes loss function $l_\mathrm{B}: D \to \mathbb{R}$ can be defined as

$$l_\mathrm{B}(d) = \int_N l(d, z) f(z) \mathrm{d}z. \tag{13}$$

Every element $d_\mathrm{B} \in D$ such that

$$l_\mathrm{B}(d_\mathrm{B}) = \min_{d \in D} l_\mathrm{B}(d) \tag{14}$$

is called a Bayes decision, and the procedure of its calculation – a Bayes decision rule. The Bayes decision is chosen in such a way, therefore, so as to minimize the expected value of losses following the decision d.

In this paper, the Bayes decision rule will be applied to the Bayes approach of the point estimation task. Namely, let the possible values of the parameters under consideration constitute the set of nature states Z with their distribution density f, while the set where we search for the values of the estimators will be treated as the set of possible decisions D. Let also be given the loss function l, describing losses resulting from estimation errors. When the above-presented Bayes decision rule has been applied to such a task, the Bayes decision becomes the desired Bayes estimator value.

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 conditioning 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 conditioning variable, respectively. Finally, let $w^* \in \mathbb{R}^{n_W}$ denote any fixed conditioning value. The goal is to calculate the estimator of this parameters' vector, denoted by $\hat{y}_{w^*} \in \mathbb{R}^{n_Y}$, 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 described in Section 2.2 will be used. Let $N = D = \mathbb{R}^{n_Y}$ naturally. For 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. It is worth noting however that in practical cases the correlations of estimation errors do not necessarily occur between all parameters; so the form can be significantly simplified with respect to the general one, depending on the specific conditions of the task considered.

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 two-dimensional loss function is assumed in a quadratic and asymmetrical form

$$l\left(\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{cases} a_l(\hat{y}_1 - y_1)^2 + a_{ld}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \qquad \text{if } \hat{y}_1 - y_1 \leq 0 \text{ and } \hat{y}_2 - y_2 \leq 0, \\ a_r(\hat{y}_1 - y_1)^2 + a_{rd}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \qquad \text{if } \hat{y}_1 - y_1 \geq 0 \text{ and } \hat{y}_2 - y_2 \leq 0, \\ a_l(\hat{y}_1 - y_1)^2 + a_{lu}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \qquad \text{if } \hat{y}_1 - y_1 \leq 0 \text{ and } \hat{y}_2 - y_2 \geq 0, \\ a_r(\hat{y}_1 - y_1)^2 + a_{ru}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \qquad \text{if } \hat{y}_1 - y_1 \geq 0 \text{ and } \hat{y}_2 - y_2 \geq 0, \end{cases} \tag{15}$$

where $a_l, a_r, a_u, a_d > 0$, $a_{ld}, a_{ru} \geq 0$ and $a_{lu}, a_{rd} \leq 0$. The coefficient a_l represents losses arising from underestimation of the first parameter (note that it is multiplied by the factor $(\hat{y}_1 - y_1)^2$ relating to the estimation error of the first parameter, and also that it occurs in these options where the condition $\hat{y}_1 - y_1 \leq 0$ showing its underestimation is obligatory). Similarly the coefficient a_r concerns losses resulting from overestimation of this parameter (it is multiplied by $(\hat{y}_1 - y_1)^2$ and appears when $\hat{y}_1 - y_1 \geq 0$, which signifies overestimation). Analogically a_d and a_u are associated with losses caused by under- and overestimation of the second parameter. In turn, the coefficients a_{ld}, a_{ru}, a_{lu}, a_{rd} represent the correlation of errors in estimation for both parameters. Thus, the coefficient a_{ld} characterizes additional losses resulting from underestimation of both parameters, a_{ru} from their simultaneous overestimation, whereas a_{lu} from underestimation of the first and overestimation of the second, and a_{rd} conversely – from overestimation of the first and underestimation of the second. It is also worth noting that in the case $a_{ld} = a_{rd} = a_{lu} = a_{ru} = 0$, which means that errors resulting from the estimation of both parameters are not correlated, the problem is reduced to two separate one-dimensional quadratic tasks considered in Section 3.2 of [8].

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

$$f_{Y|W=w^*}(y_1, y_2) = f_{Y_1|W=w^*}(y_1) f_{Y_2|W=w^*}(y_2). \tag{16}$$

Let also the functions $f_{Y_1|W=w^*}$ and $f_{Y_2|W=w^*}$ be continuous, and such that

$$\int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 < \infty, \tag{17}$$

$$\int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 < \infty. \tag{18}$$

The Bayes loss function (13) for the losses described by the formula (15) is therefore given as

$$\begin{aligned}
l_{\mathrm{B}}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) &= a_r \int_{-\infty}^{\hat{y}_1} (\hat{y}_1 - y_1)^2 f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{\infty} f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_l \int_{\hat{y}_1}^{\infty} (\hat{y}_1 - y_1)^2 f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{\infty} f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_{rd} \int_{-\infty}^{\hat{y}_1} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_{ld} \int_{\hat{y}_1}^{\infty} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_{lu} \int_{\hat{y}_1}^{\infty} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_{ru} \int_{-\infty}^{\hat{y}_1} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_d \int_{-\infty}^{\infty} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2)^2 f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ a_u \int_{-\infty}^{\infty} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 \int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2)^2 f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2.
\end{aligned} \tag{19}$$

Taking into account the assumptions made earlier concerning continuity of the functions $f_{Y_1|W=w^*}$ and $f_{Y_2|W=w^*}$, the partial derivatives of the function l_{B} exist and are

$$\begin{aligned}
\frac{\partial l_{\mathrm{B}}}{\partial \hat{y}_1}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) &= \left(a_{ld} \int_{\hat{y}_1}^{\infty} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 + a_{rd} \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1\right) \\
&\times \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_1|W=w^*}(y_2)\, \mathrm{d}y_2 + \left(a_{lu} \int_{\hat{y}_1}^{\infty} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1\right. \\
&\left. + a_{ru} \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1\right) \int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\, \mathrm{d}y_2 \\
&+ 2a_l \int_{\hat{y}_1}^{\infty} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1 + 2a_r \int_{-\infty}^{\hat{y}_1} (\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\, \mathrm{d}y_1,
\end{aligned} \tag{20}$$

$$\frac{\partial l_{\mathrm{B}}}{\partial \hat{y}_2}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) = \left(a_{ld}\int_{\hat{y}_2}^{\infty} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + a_{lu}\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right)$$
$$\times\int_{\hat{x}_1}^{\infty}(\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + \left(a_{rd}\int_{\hat{y}_2}^{\infty} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right. \tag{21}$$
$$\left. + a_{ru}\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right)\int_{-\infty}^{\hat{y}_1}(\hat{y}_1 - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1$$
$$+ 2a_d\int_{\hat{y}_2}^{\infty}(\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + 2a_u\int_{-\infty}^{\hat{y}_2}(\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2.$$

Making use of the additivity of an integral with respect to an integration set, and equating them to zero one obtains the equations which are a necessary condition for the extreme of this function:

$$\int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\left[(a_{ru} - a_{rd} - a_{lu} + a_{ld})\left(\hat{y}_2\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right.\right.$$
$$\left.\left. - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right) + (a_{rd} - a_{ld})\left(\hat{y}_2 - \int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right)\right]$$
$$+ 2a_l\left(\hat{y}_1 - \int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right) + 2(a_r - a_l)\left(\hat{y}_1\int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right.$$
$$\left. - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right) + a_{ld}\left(\hat{y}_2 - \int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right)$$
$$+ (a_{lu} - a_{ld})\left(\hat{y}_2\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right) = 0, \tag{22}$$

$$\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\left[(a_{ru} - a_{rd} - a_{lu} + a_{ld})\left(\hat{y}_1\int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right.\right.$$
$$\left.\left. - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right) + (a_{lu} - a_{ld})\left(\hat{y}_1 - \int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\right]$$
$$+ 2a_d\left(\hat{y}_2 - \int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right) + 2(a_u - a_d)\left(\hat{y}_2\int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right.$$
$$\left. - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\right) + a_{ld}\left(\hat{y}_1 - \int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)$$
$$+ (a_{rd} - a_{ld})\left(\hat{y}_1\int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right) = 0. \tag{23}$$

Appendix 1 shows that a solution exists, is unique and constitutes a global minimum of the function l_{B}. Calculating its value in the general case, however, is not an easy task, although if estimation of the densities present above is reached using statistical kernel estimators described in Section 2.1, 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 $U_{1,i} : \mathbb{R} \to \mathbb{R}$, $U_{2,i} : \mathbb{R} \to \mathbb{R}$, $V_{1,i} : \mathbb{R} \to \mathbb{R}$ and $V_{2,i} : \mathbb{R} \to \mathbb{R}$, given as

$$U_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} K\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathrm{d}y_1, \tag{24}$$

$$U_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} K\left(\frac{y_2 - y_{i,2}}{h_2}\right) \mathrm{d}y_2, \tag{25}$$

$$V_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} y_1 \, K\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathrm{d}y_1, \tag{26}$$

$$V_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} y_2 \, K\left(\frac{y_2 - y_{i,2}}{h_2}\right) \mathrm{d}y_2. \tag{27}$$

Norm also the conditioning parameters d_i by introducing the positive values

$$d_i^* = \frac{d_i}{\sum_{i=1}^m d_i} \quad \text{for } i = 1, 2, \ldots, m; \tag{28}$$

note that

$$\sum_{i=1}^m d_i^* = 1. \tag{29}$$

After entering the above notations and allowing for the form of the kernel estimator of conditioning random variable (12), criterions (25) and (26) can be given in the equivalent form

$$\begin{aligned}
&\sum_{i=1}^m d_i^* \, U_{1,i}(\hat{y}_1) \Bigg[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) \\
&\quad + (a_{rd} - a_{ld}) \left(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2} \right) \Bigg] + 2a_l \left(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1} \right) + 2(a_r - a_l) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) \\
&\quad - V_{1,i}(\hat{y}_1)) + a_{ld} \left(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2} \right) + (a_{lu} - a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) = 0,
\end{aligned} \tag{30}$$

$$\begin{aligned}
&\sum_{i=1}^m d_i^* \, U_{2,i}(\hat{y}_2) \Bigg[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) \\
&\quad + (a_{lu} - a_{ld}) \left(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1} \right) \Bigg] + 2a_d \left(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2} \right) + 2(a_u - a_d) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) \\
&\quad - V_{2,i}(\hat{y}_2)) + a_{ld} \left(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1} \right) + (a_{rd} - a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) = 0.
\end{aligned} \tag{31}$$

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 as

$$\begin{aligned}
\frac{\partial L_1(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_1} &= \sum_{i=1}^m d_i^* \frac{1}{h_1 s_{i,1}} K\left(\frac{\hat{y}_1 - y_{i,1}}{h_1 s_{i,1}}\right) \Bigg[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) \\
&\quad + (a_{rd} - a_{ld})(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2}) \Bigg] + 2(a_r - a_l) \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) + 2a_l,
\end{aligned} \tag{32}$$

$$\frac{\partial L_1(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2} = \sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+(a_{rd}-a_{ld})\right] + (a_{lu}-a_{ld})\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+a_{ld}, \tag{33}$$

$$\frac{\partial L_2(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_1} = \sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)+(a_{lu}-a_{ld})\right] + (a_{rd}-a_{ld})\sum_{i=1}^{m} d_i^* U_{1,i}(\hat{y}_1)+a_{ld}, \tag{34}$$

$$\frac{\partial L_2(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2} = \sum_{i=1}^{m} d_i^* \frac{1}{h_2 s_{i,2}} K\left(\frac{\hat{y}_2-y_{i,2}}{h_2 s_{i,2}}\right)\left[(a_{ru}-a_{rd}-a_{lu}+a_{ld})\sum_{i=1}^{m} d_i^*(\hat{y}_1 U_{1,i}(\hat{y}_1)-V_{1,i}(\hat{y}_1)) + (a_{lu}-a_{ld})\left(\hat{y}_1-\sum_{i=1}^{m} d_i^* y_{i,1}\right)\right] + 2(a_u-a_d)\sum_{i=1}^{m} d_i^* U_{2,i}(\hat{y}_2)+2a_d. \tag{35}$$

Then the solution of Equations (22) and (23) can be calculated through Newton's multidimensional algorithm [15] 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

$$\hat{y}_{0,1} = \frac{\sum_{i=1}^{m} d_i y_{i,1}}{\sum_{i=1}^{m} d_i}, \tag{36}$$

$$\hat{y}_{0,2} = \frac{\sum_{i=1}^{m} d_i y_{i,2}}{\sum_{i=1}^{m} d_i}, \tag{37}$$

$$\begin{bmatrix}\hat{y}_{j+1,1}\\ \hat{y}_{j+1,2}\end{bmatrix} = \begin{bmatrix}\hat{y}_{j,1}\\ \hat{y}_{j,2}\end{bmatrix} - \begin{bmatrix}\dfrac{\partial L_1(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_1} & \dfrac{\partial L_1(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_2}\\ \dfrac{\partial L_2(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_1} & \dfrac{\partial L_2(\hat{y}_{j,1},\hat{y}_{j,2})}{\partial \hat{y}_2}\end{bmatrix}^{-1} \begin{bmatrix}L_1(\hat{y}_{j,1},\hat{y}_{j,2})\\ L_2(\hat{y}_{j,1},\hat{y}_{j,2})\end{bmatrix} \quad \text{for } j=0,1,\ldots, \tag{38}$$

while the quantities in the above dependencies are given by Equations (30)–(35), whereas a stop condition takes the form of the conjunction of the following inequalities:

$$|\hat{y}_{j,1}-\hat{y}_{j-1,1}| \le 0{,}01\,\hat{\sigma}_1, \tag{39}$$

$$|\hat{y}_{j,2}-\hat{y}_{j-1,2}| \le 0{,}01\,\hat{\sigma}_2, \tag{40}$$

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

4. Kernel used

As mentioned, the important positive feature of kernel estimators constitutes a possibility to choose the kernel form with regard to the demands of the practical task being worked out. With respect to the previous considerations, the following four requirements should be formulated:

- continuity and positivity of the function K;
- the finite first moment of the above function, i.e. $\int_{-\infty}^{\infty} xK(x)\,\mathrm{d}x < \infty$;
- the function $I : \mathbb{R} \to \mathbb{R}$ defined as $I(x) = \int_{-\infty}^{x} K(y)\,\mathrm{d}y$ should be expressed by a relatively simple analytical formula;
- similar to the above with respect to the real function $J : \mathbb{R} \to \mathbb{R}$ given by $J(x) = \int_{-\infty}^{x} yK(y)\,\mathrm{d}y$.

Because of the above conditions, the Cauchy kernel

$$K(x) = \frac{2}{\pi}\frac{1}{(1+x^2)^2} \tag{41}$$

can be proposed. Then

$$U_i(\hat{y}) = \frac{1}{\pi}\mathrm{arctg}\left(\frac{\hat{y}-y_i}{h}\right) + \frac{\dfrac{\hat{y}-y_i}{h}}{\pi\left[1+\left(\dfrac{\hat{y}-y_i}{h}\right)^2\right]} + \frac{1}{2}, \tag{42}$$

$$V_i(\hat{y}) = y_i\left[\frac{1}{\pi}\mathrm{arctg}\left(\frac{\hat{y}-y_i}{h}\right) + \frac{\dfrac{\hat{y}-y_i}{h}}{\pi\left[1+\left(\dfrac{\hat{y}-y_i}{h}\right)^2\right]} + \frac{1}{2}\right] - \frac{h}{\pi\left[1+\left(\dfrac{\hat{y}-y_i}{h}\right)^2\right]}, \tag{43}$$

while if one applies the plug-in method recommended here, the constants there amount to $\int_{\mathbb{R}} x^2K(x)\,\mathrm{d}x = 1$ and $\int_{\mathbb{R}} K(x)^2\,\mathrm{d}x = 5/4\pi$.

5. Numerical testing

The correctness of the investigated algorithm was comprehensively examined both with illustrative artificial data obtained from a generator (acquired results are presented below) and in experimental research concerning practical tasks, described in Section 7.

Assume – for transparency of the results interpretation – that $n_Y = 2$ and $n_W = 1$, and let the tested random variable $X = \begin{bmatrix} Y \\ W \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ W \end{bmatrix}$ has distribution being the sum of five Gauss factors with expected values, covariance matrixes and shares, respectively,

$$E_1 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathrm{Cov}_1 = \begin{bmatrix} 1 & -0.49 & 0.7 \\ -0.49 & 1 & -0.7 \\ 0.7 & -0.7 & 1 \end{bmatrix}, \quad 30\%, \tag{44}$$

$$E_2 = \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathrm{Cov}_2 = \begin{bmatrix} 1 & -0.49 & 0.7 \\ -0.49 & 1 & -0.7 \\ 0.7 & -0.7 & 1 \end{bmatrix}, \quad 20\%, \tag{45}$$

$$E_3 = \begin{bmatrix} -2 \\ -2 \\ 0 \end{bmatrix}, \quad \mathrm{Cov}_3 = \begin{bmatrix} 1 & -0.49 & 0.7 \\ -0.49 & 1 & -0.7 \\ 0.7 & -0.7 & 1 \end{bmatrix}, \quad 15\%, \tag{46}$$

$$E_4 = \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix}, \quad \mathrm{Cov}_4 = \begin{bmatrix} 1 & -0.49 & 0.7 \\ -0.49 & 1 & -0.7 \\ 0.7 & -0.7 & 1 \end{bmatrix}, \qquad 15\%, \tag{47}$$

$$E_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathrm{Cov}_5 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad 20\%. \tag{48}$$

Comments on conditionally independent random variables can be found in Appendix 2. In the case of factors (44)–(47), the conditioning variable W is positively correlated with the first coordinate of the describing variable Y, i.e. with Y_1, and negatively with the second coordinate Y_2. Factor (48), with uncorrelated coordinates, acts as an additional – apart from the natural uncertainty of the variables Y and W – disturbance. The expected values of variables Y_1, Y_2 and W as well as their standard deviations are

$$E_{Y_1} = 0.2, \quad \sigma_{Y_1} \cong 2.2, \tag{49}$$

$$E_{Y_2} = 0.4, \quad \sigma_{Y_2} \cong 2.2, \tag{50}$$

$$E_W = 0, \quad \sigma_W = 1. \tag{51}$$

The results acquired with the algorithm described in Sections 3 and 4, for $w^* = 0$, $w^* = 1$, $w^* = 2$, so for the modal value of the conditioning random variable as well as at the first and second standard deviation, are presented in Tables 1–3, respectively. Each of their cells shows the obtained values of the estimator, calculated on the basis of 100 tests and recorded in the classic formula: 'mean value $\pm$ standard deviation'. The symbol ∞ denotes the analytically achieved theoretical value.

In Tables 1–3 the results have been shaded where the mean estimation error is greater than 10% of the standard deviation of the describing variable – i.e. 0.22 for both coordinates Y_1 and Y_2 – or where the standard deviation of the estimation error is greater than 20% of the standard deviation of the describing variable – i.e. 0.44 for Y_1 and Y_2 (see formulas (49)–(51)). One can note that the remaining (unshaded) results are for samples of sizes from 50 to 100 for $w^* = 0$, 100 to 200 for $w^* = 1$, 200 to 500 for $w^* = 2$. Taking into account that the distribution of the describing random variable Y is four-modal, the existence of an additional disturbance represented by factor (48), and the complex multidimensional character of the task, the need for such size seems reasonable in practice. In particular, estimation at the second standard deviation of the conditioning variable (Table 3) appears to be an especially difficult problem, which justifies the need for a sample size of at least 500.

Analysis of particular columns of Tables 1–3 will now be discussed.

The first three columns serve an auxiliary purpose. Above all, the first constitutes a reference point. Through the assumption $a_{ld} = a_{rd} = a_{lu} = a_{ru} = 0$, it means that the estimation errors of both parameters are not correlated, and thus the problem is reduced to two separate one-dimensional quadratic tasks worked out in Section 3.2 of [8]. Additionally, the first case $a_l = a_r = a_d = a_u = 1$ shows the problem with symmetrical losses.

In the second column, errors of estimation are still uncorrelated, while the coefficient a_l value is three times larger. According to the form of loss function (15), it means that underestimations of the first parameters result in greater losses than overestimations. In consequence, the calculated value of the first coordinate of the estimator is greater than that in the first column, which lowers the probability of underestimations unfavourable here. The value of the second coordinate of the estimators remains – due to the lack of changes in the coefficients a_d and a_u – unaltered.

Table 1. Results of simulations for $w^* = 0$.

m	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 10.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$
∞	0.200 0.400	1.146 0.400	1.899 0.400
50	0.219 ± 0.394 0.520 ± 0.413	1.196 ± 0.345 0.520 ± 0.413	2.049 ± 0.322 0.520 ± 0.413
100	0.184 ± 0.305 0.426 ± 0.284	1.157 ± 0.261 0.426 ± 0.284	1.982 ± 0.239 0.426 ± 0.284
200	0.188 ± 0.216 0.433 ± 0.228	1.153 ± 0.189 0.433 ± 0.228	1.961 ± 0.180 0.433 ± 0.228
500	0.211 ± 0.166 0.415 ± 0.171	1.165 ± 0.142 0.415 ± 0.171	1.951 ± 0.130 0.415 ± 0.171
1000	0.210 ± 0.119 0.409 ± 0.127	1.158 ± 0.100 0.409 ± 0.127	1.930 ± 0.092 0.409 ± 0.127
2000	0.204 ± 0.092 0.405 ± 0.094	1.156 ± 0.076 0.405 ± 0.094	1.925 ± 0.070 0.405 ± 0.094
5000	0.210 ± 0.063 0.402 ± 0.064	1.159 ± 0.054 0.402 ± 0.064	1.923 ± 0.048 0.402 ± 0.064
10,000	0.204 ± 0.051 0.406 ± 0.050	1.153 ± 0.043 0.406 ± 0.050	1.915 ± 0.036 0.406 ± 0.050

$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 3.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 10.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 3.0$ $a_{lu} = -1.0, a_{ru} = 3.0$ $a_{ld} = 3.0, a_{rd} = -10.0$ $a_u = 3.0, a_d = 10.0$	m
0.670 0.957	1.019 1.518	−0.142 1.573	∞
0.675 ± 0.440 1.051 ± 0.442	1.042 ± 0.519 1.579 ± 0.460	−0.113 ± 0.459 1.672 ± 0.336	50
0.650 ± 0.351 0.964 ± 0.317	1.018 ± 0.428 1.516 ± 0.343	−0.165 ± 0.367 1.602 ± 0.222	100
0.652 ± 0.249 0.981 ± 0.257	0.997 ± 0.311 1.551 ± 0.278	−0.157 ± 0.266 1.605 ± 0.179	200
0.678 ± 0.192 0.963 ± 0.193	1.014 ± 0.237 1.545 ± 0.200	−0.131 ± 0.210 1.586 ± 0.141	500
0.685 ± 0.141 0.956 ± 0.146	1.024 ± 0.183 1.540 ± 0.156	−0.130 ± 0.151 1.584 ± 0.104	1000
0.677 ± 0.109 0.958 ± 0.109	1.007 ± 0.149 1.553 ± 0.124	−0.139 ± 0.117 1.581 ± 0.077	2000
0.683 ± 0.074 0.955 ± 0.074	1.008 ± 0.101 1.553 ± 0.083	−0.131 ± 0.082 1.578 ± 0.052	5000
0.675 ± 0.061 0.961 ± 0.059	0.996 ± 0.085 1.562 ± 0.064	−0.137 ± 0.066 1.579 ± 0.038	10,000

In the third column, thanks to the next increase in the value of the coefficient a_l the above effect is intensified.

In interpreting the fourth column, where losses arising from over- and underestimation are now correlated, it is worth comparing with the basic first column. In the fourth column, the value of the

Table 2. Results of simulations for $w^* = 1$.

m	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 10.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$
∞	0.760 −0.160	1.738 −0.160	2.512 −0.160
50	0.649 ± 0.537 -0.091 ± 0.458	1.657 ± 0.480 -0.091 ± 0.458	2.574 ± 0.451 -0.091 ± 0.458
100	0.726 ± 0.348 -0.171 ± 0.409	1.718 ± 0.298 -0.171 ± 0.409	2.578 ± 0.267 -0.171 ± 0.409
200	0.751 ± 0.276 -0.183 ± 0.324	1.729 ± 0.243 -0.183 ± 0.324	2.560 ± 0.209 -0.183 ± 0.324
500	0.775 ± 0.229 -0.171 ± 0.214	1.745 ± 0.201 -0.171 ± 0.214	2.552 ± 0.169 -0.171 ± 0.214
1000	0.756 ± 0.153 -0.178 ± 0.166	1.732 ± 0.128 -0.178 ± 0.166	2.532 ± 0.107 -0.178 ± 0.166
2000	0.766 ± 0.120 -0.168 ± 0.122	1.742 ± 0.099 -0.168 ± 0.122	2.535 ± 0.080 -0.168 ± 0.122
5000	0.765 ± 0.085 -0.171 ± 0.077	1.741 ± 0.070 -0.171 ± 0.077	2.529 ± 0.055 -0.171 ± 0.077
10,000	0.766 ± 0.065 -0.162 ± 0.052	1.743 ± 0.055 -0.162 ± 0.052	2.527 ± 0.044 -0.162 ± 0.052

$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 3.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 10.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 3.0$ $a_{lu} = -1.0, a_{ru} = 3.0$ $a_{ld} = 3.0, a_{rd} = -10.0$ $a_u = 3.0, a_d = 10.0$	m
1.187 0.433	1.478 1.020	0.394 0.982	∞
1.081 ± 0.590 0.460 ± 0.488	1.401 ± 0.657 1.037 ± 0.503	0.286 ± 0.629 1.077 ± 0.385	50
1.175 ± 0.394 0.378 ± 0.437	1.498 ± 0.479 0.966 ± 0.469	0.362 ± 0.417 1.006 ± 0.337	100
1.202 ± 0.320 0.368 ± 0.359	1.529 ± 0.400 0.956 ± 0.367	0.390 ± 0.330 0.986 ± 0.259	200
1.227 ± 0.266 0.387 ± 0.238	1.540 ± 0.331 0.984 ± 0.252	0.420 ± 0.276 0.990 ± 0.173	500
1.202 ± 0.179 0.394 ± 0.177	1.502 ± 0.225 1.006 ± 0.176	0.394 ± 0.189 0.984 ± 0.137	1000
1.207 ± 0.138 0.408 ± 0.132	1.496 ± 0.169 1.024 ± 0.127	0.405 ± 0.156 0.987 ± 0.100	2000
1.206 ± 0.100 0.408 ± 0.088	1.495 ± 0.126 1.022 ± 0.091	0.401 ± 0.109 0.984 ± 0.058	5000
1.202 ± 0.074 0.423 ± 0.059	1.481 ± 0.089 1.038 ± 0.057	0.403 ± 0.083 0.987 ± 0.038	10,000

coefficient a_{ld} grows, therefore – as results from formula (15) – simultaneous underestimation of both parameters is a disadvantage. To avoid this, the values of both coordinates of the estimators are greater than those from the first column.

Table 3. Results of simulations for $w^* = 2$.

m	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 10.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$
∞	1.320 −0.720	2.341 −0.720	3.147 −0.720
50	1.026 ± 0.860 -0.421 ± 0.775	2.006 ± 0.834 -0.421 ± 0.775	2.937 ± 0.815 -0.421 ± 0.775
100	1.063 ± 0.682 -0.461 ± 0.613	2.070 ± 0.641 -0.461 ± 0.613	3.012 ± 0.600 -0.461 ± 0.613
200	1.142 ± 0.427 -0.542 ± 0.417	2.146 ± 0.436 -0.542 ± 0.443	3.045 ± 0.450 -0.542 ± 0.471
500	1.209 ± 0.322 -0.610 ± 0.350	2.227 ± 0.305 -0.610 ± 0.350	3.105 ± 0.287 -0.610 ± 0.350
1000	1.231 ± 0.226 -0.604 ± 0.251	2.252 ± 0.200 -0.604 ± 0.251	3.116 ± 0.176 -0.604 ± 0.251
2000	1.251 ± 0.171 -0.605 ± 0.175	2.271 ± 0.149 -0.605 ± 0.175	3.121 ± 0.130 -0.605 ± 0.175
5000	1.259 ± 0.131 -0.666 ± 0.110	2.282 ± 0.113 -0.666 ± 0.110	3.123 ± 0.098 -0.666 ± 0.110
10,000	1.265 ± 0.094 -0.687 ± 0.093	2.288 ± 0.082 -0.687 ± 0.093	3.122 ± 0.071 -0.687 ± 0.093

$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 3.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 10.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 3.0$ $a_{lu} = -1.0, a_{ru} = 3.0$ $a_{ld} = 3.0, a_{rd} = -10.0$ $a_u = 3.0, a_d = 10.0$	m
1.718 −0.009	1.997 0.448	0.926 0.406	∞
1.440 ± 0.937 0.083 ± 0.824	1.756 ± 1.001 0.592 ± 0.886	0.723 ± 0.909 0.666 ± 0.698	50
1.458 ± 0.758 0.068 ± 0.646	1.751 ± 0.849 0.608 ± 0.655	0.718 ± 0.766 0.634 ± 0.500	100
1.551 ± 0.563 0.003 ± 0.490	1.855 ± 0.624 0.567 ± 0.489	0.779 ± 0.591 0.573 ± 0.358	200
1.630 ± 0.368 -0.034 ± 0.367	1.938 ± 0.440 0.555 ± 0.346	0.832 ± 0.381 0.529 ± 0.262	500
1.644 ± 0.256 -0.010 ± 0.265	1.940 ± 0.320 0.581 ± 0.248	0.849 ± 0.272 0.531 ± 0.193	1000
1.659 ± 0.194 -0.002 ± 0.185	1.953 ± 0.243 0.588 ± 0.177	0.868 ± 0.210 0.526 ± 0.141	2000
1.668 ± 0.147 -0.055 ± 0.118	1.952 ± 0.180 0.547 ± 0.107	0.867 ± 0.165 0.471 ± 0.088	5000
1.674 ± 0.109 -0.075 ± 0.097	1.958 ± 0.141 0.529 ± 0.087	0.871 ± 0.118 0.450 ± 0.073	10,000

In the fifth column, the coefficient a_{ld} value has been increased even more, which intensifies the above phenomenon.

And lastly, in the final sixth column, also with correlated losses, all coefficients have been modified with respect to the basic case from the first column, with a_d and a_{rd} to the greatest degree. The former, a_d, according to the form of loss function (15), indicates greater losses resulting from the underestimation of the second parameter, the latter a_{rd} those from the overestimation of the first parameter and – again – underestimation of the second. As a result, the value of the first coordinate of the estimator is less than that in the first column (the influence of the coefficient a_d), while due to the two factors (both coefficients a_d and a_{rd}) the value of its second coordinate grows even greater still.

The results in Tables 1–3 will now be collated, comparing the values obtained for the different conditioning values $w^* = 0$, $w^* = 1$ and $w^* = 2$, respectively. The results for $w^* = -1$ and $w^* = -2$ have been omitted here as symmetrical with respect to $w^* = 1$ and $w^* = 2$. As can be seen from formulas (44)–(48), the first coordinate of the describing variable Y_1 is positively correlated to the conditioning variable W and the second Y_2 negatively. Therefore, comparing any column from Tables 1–3, one can see that the estimators' values for $w^* = 1$ are greater in the case of the first coordinate and smaller in that of the second than for $w^* = 0$. For $w^* = 2$ the differences become even more distinct. One may thus conclude that by entering the proper (e.g. current) values of conditioning variables, the values of the estimated parameters, and consequently the model used, can be made more precise in practical tasks.

In many applications, for given values of a conditioning factor, there may not be a sufficient number of measurements in the database in possession for reliable statistical inference. For this reason, methods considering conditioning factors should be robust with respect to insufficient number or even lack of measurements in some areas of potential values of such factors. In Table 4, the results are shown for the case $w^* = 1$ obtained only for those elements of random sample (6), for which $w_i \notin [0.9, 1.1]$. Comparing them with the contents of Table 1, one may note that, despite a complete lack of data in the 'belt' of width 0.2, or 10% of the value of the standard deviation of the conditioning variable, the results were not significantly worse. This fact shows well the robustness of the proposed algorithm with respect to a small amount, or even lack of data in certain areas of potential values of conditioning factors. This effect is fundamentally justified – it stems from averaging properties of kernel estimators.

When interpreting the results of Tables 1–4, one can see that the estimating procedure worked out in this paper for the case of correlated losses resulting from over- and underestimations (see lower columns 4–6) is obviously more difficult than for no such correlation (see auxiliary upper columns 1–3). It should also be added that, where there is no correlation of losses, even better effects can be obtained by applying the algorithm designated for a single parameter, presented in Section 2 of [8].

Comparing the fourth (where $a_{ld} = 3$) and fifth (where $a_{ld} = 10$) columns, it can be noted that estimation is easier in the first case, with closer values of coefficients of the loss function (15). It is confirmed by the results in column six, where the values are admittedly overall greater but closer to one another. It is worth pointing out that the very form of loss function (15) shows that absolute values of the coefficients have no influence on the results but rather the relations (proportions) between them.

Finally, it is worth noting that in any case shown in Tables 1–4, as the sample size increased, the obtained parameter value converged to the theoretical, and the standard deviation to zero. Such asymptotical features are of fundamental significance from an applicational point of view, as they prove that it is possible to obtain any precision wished, although this requires the assurance of a sufficient random sample size and computational capability of computer system used. In practice, therefore, the necessity of the right compromise between these quantities is called for.

The above corollaries have been successfully proved in numerous obtained results of simulations, also for a multidimensional conditioning variable, and multimodal, asymmetrical and complex distributions of the variables Y and W, as well as those including additional aspects, e.g. a bounded support or data censored.

Table 4. Results of simulations for $w^* = 1$ when lacking data with conditioning variable values from the interval [0.9, 1.1].

m	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 10.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 0.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$
∞	0.760 −0.160	1.738 −0.160	2.512 −0.1600
50	0.588 ± 0.562 −0.164 ± 0.541	1.582 ± 0.500 −0.164 ± 0.541	2.500 ± 0.477 −0.164 ± 0.541
100	0.713 ± 0.408 −0.200 ± 0.409	1.697 ± 0.347 −0.200 ± 0.409	2.564 ± 0.318 −0.200 ± 0.409
200	0.745 ± 0.320 −0.174 ± 0.324	1.727 ± 0.273 −0.174 ± 0.324	2.573 ± 0.249 −0.174 ± 0.324
500	0.758 ± 0.224 −0.155 ± 0.215	1.734 ± 0.189 −0.155 ± 0.215	2.553 ± 0.165 −0.155 ± 0.215
1000	0.743 ± 0.170 −0.172 ± 0.156	1.722 ± 0.142 −0.172 ± 0.156	2.532 ± 0.125 −0.172 ± 0.156
2000	0.760 ± 0.121 −0.171 ± 0.127	1.737 ± 0.099 −0.171 ± 0.127	2.537 ± 0.089 −0.171 ± 0.127
5000	0.773 ± 0.104 −0.174 ± 0.086	1.747 ± 0.087 −0.174 ± 0.086	2.536 ± 0.076 −0.174 ± 0.086
10,000	0.773 ± 0.078 −0.158 ± 0.071	1.748 ± 0.064 −0.158 ± 0.071	2.532 ± 0.051 −0.158 ± 0.071

$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 3.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 1.0, a_r = 1.0$ $a_{lu} = 0.0, a_{ru} = 0.0$ $a_{ld} = 10.0, a_{rd} = 0.0$ $a_u = 1.0, a_d = 1.0$	$a_l = 3.0, a_r = 3.0$ $a_{lu} = -1.0, a_{ru} = 3.0$ $a_{ld} = 3.0, a_{rd} = -10.0$ $a_u = 3.0, a_d = 10.0$	m
1.187 0.433	1.478 1.020	0.394 0.982	∞
1.033 ± 0.609 0.361 ± 0.566	1.387 ± 0.693 0.860 ± 0.626	0.220 ± 0.656 1.011 ± 0.462	50
1.166 ± 0.432 0.334 ± 0.426	1.506 ± 0.545 0.843 ± 0.515	0.350 ± 0.430 0.975 ± 0.328	100
1.194 ± 0.366 0.374 ± 0.365	1.509 ± 0.433 0.907 ± 0.423	0.386 ± 0.375 0.991 ± 0.255	200
1.202 ± 0.263 0.406 ± 0.240	1.491 ± 0.365 0.984 ± 0.282	0.404 ± 0.269 1.001 ± 0.167	500
1.186 ± 0.199 0.400 ± 0.171	1.433 ± 0.253 1.011 ± 0.182	0.381 ± 0.207 0.988 ± 0.123	1000
1.201 ± 0.146 0.403 ± 0.141	1.453 ± 0.190 1.012 ± 0.138	0.400 ± 0.151 0.984 ± 0.103	2000
1.214 ± 0.125 0.402 ± 0.100	1.473 ± 0.144 1.007 ± 0.104	0.412 ± 0.126 0.979 ± 0.069	5000
1.208 ± 0.092 0.424 ± 0.079	1.479 ± 0.107 1.030 ± 0.067	0.411 ± 0.097 0.988 ± 0.055	10,000

6. Additional comments

The procedure presented in this paper has been given in its basic form, easier to implement and computationally more convenient. A clear interpretation means it is possible to make individual modifications and generalizations, which may be useful in particular atypical tasks.

Above all this allows the inclusion of conditioning factors other than continuous (real). Similarly to the kernel estimation definition (2) formulated for continuous random variables, one can construct kernel estimators for discrete, multivalued (in particular binary) and categorized (nominal and ordered) variables, as well as any of their compositions, especially with continuous variables. The literature concerning this subject is quite broad and varied. For the first case, it is worth quoting the articles [16,17] and for the second the classic monographs [11,Section 3.1.8;13,Section 6.1.4] as well as the paper [18]. Issues connected with categorical variables can be found in the publications.[19–21] After introducing discrete, multivalued (also binary) and/or categorized variable to the algorithm worked out here, it undergoes practically no changes, apart from technical ones resulting from calculational differences. This property particularly should be underlined considering the modern data analysis tasks, which more and more often take advantage of the many different configurations for particular types of attributes.

Newton's method ((36)–(40)) can also be applied in numerous mutations available in the literature, in particular those lessening the number of iterations as well as extending the convergence area. A broad review of concepts available on this subject can be found in the monographs.[22,23] It must, however, be underlined that in the research undertaken, the problem of no convergence for sizes of sample guaranteeing a satisfactory estimation quality (see Section 5) did not arise, largely due to the choice of the starting point in form (36)–(37).

The quadratic loss function (15) can be also generalized to a polynomial asymmetrical form

$$l\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}, \begin{bmatrix}y_1\\ y_2\end{bmatrix}\right) = \begin{cases} a_l|\hat{y}_1 - y_1|^q + a_{ld}|\hat{y}_1 - y_1|^{q/2}|\hat{y}_2 - y_2|^{q/2} + a_d|\hat{y}_2 - y_2|^q \\ \qquad \text{if } \hat{y}_1 - y_1 \le 0 \text{ and } \hat{y}_2 - y_2 \le 0, \\ a_r|\hat{y}_1 - y_1|^q + a_{rd}|\hat{y}_1 - y_1|^{q/2}|\hat{y}_2 - y_2|^{q/2} + a_d|\hat{y}_2 - y_2|^q \\ \qquad \text{if } \hat{y}_1 - y_1 \ge 0 \text{ and } \hat{y}_2 - y_2 \le 0, \\ a_l|\hat{y}_1 - y_1|^q + a_{lu}|\hat{y}_1 - y_1|^{q/2}|\hat{y}_2 - y_2|^{q/2} + a_u|\hat{y}_2 - y_2|^q \\ \qquad \text{if } \hat{y}_1 - y_1 \le 0 \text{ and } \hat{y}_2 - y_2 \ge 0, \\ a_r|\hat{y}_1 - y_1|^q + a_{ru}|\hat{y}_1 - y_1|^{q/2}|\hat{y}_2 - y_2|^{q/2} + a_u|\hat{y}_2 - y_2|^q \\ \qquad \text{if } \hat{y}_1 - y_1 \ge 0 \text{ and } \hat{y}_2 - y_2 \ge 0, \end{cases} \tag{52}$$

where $a_l, a_r, a_u, a_d > 0$ and $a_{ld}, a_{ru}, a_{lu}, a_{rd} \ge 0$; an interpretation of these parameters remains the same. The positive parameter q, introduced above, defines the degree of polynomial. Having previously investigated the quadratic case $q = 2$, consider now the related linear $q = 1$, cubic $q = 3$ and fourth order $q = 4$ (the formulas presented in this paper undergo the appropriate changes, but the idea remains analogous to that described in Section 3). Thus, the greater the degree of the polynomial, the lower the sensitivity of the calculated estimator to asymmetry of loss function (52), and so the resulting values converge to the means and medians of particular coordinates. This is due to the fact that, for a greater power q, big estimation errors have a significantly increased influence on the value of the loss function; therefore the optimal value of the estimator is defined so as to avoid just such huge elements, and consequently it is 'safest' to remain near the mean/median. In consequence, using asymmetric loss function (52) with degree q greater than four seems to be without sense in most applications. Generally, the polynomial degree should be fixed according to the conditions of the specific task, in particular based on the dependence of loss value on potential estimation errors. It is also worth paying attention to additional aspects, for example to the problem of the so-called censored data, i.e. the situation where values above and/or below a given margin are transposed to the appropriate – upper or lower – boundary value. This situation occurs – for example – with a limited range of measurements, where values exceeding this range must be reduced to the proper boundary. In this situation, it is advisable to use the lowest possible polynomial degree, which decreases the sensitivity of received results to the effects of reducing extreme values, forced by censoring.

An important question is the computational complexity of the investigated procedure. Above all, one should underline the advantage – from the practical point of view – of its having two stages. The first phase contains algorithms for calculating parameter values. The plug-in method, used to achieve the smoothing parameter h value for every $n_Y + n_W$ coordinates, and the algorithm for its modification have computational complexity $O(m^2)$, but they are carried out once at the beginning of the procedure. The second phase consists of finding a solution for criterion (22)–(23), with the aid of Newton's algorithm (36)–(40), with the calculation of the parameters d_i having the complexity $O(n_W)$, while the functions L and L' are of complexity $O(m)$, and so are linear. Newton's algorithm in investigated cases most often required 10–15 iterations. This implies a relatively short computation time for the second phase. It means that with the first phase having been carried out earlier, in most practical tasks it is possible to apply the worked-out procedure in real time. However, even without division into stages, computation time did not exceed 1 s for sample size (1) to $m = 1000$, and 1 min for $m = 10{,}000$, while computations were carried out on rudimentary equipment, without particular efforts to shorten them. This was possible to a great extent, thanks to the proper choice of kernel shape, and thanks to that, analytical forms of the functions (24)–(27) used.

7. Experimental research

The procedure worked out here also underwent verification in real applicational tasks both in an automatic control task by numerical simulation as well as by experiment in a problem from the field of medicine.

First, consider time-optimal control, consisting of bringing the state of a dynamic system to the assumed target in a minimal time.[24] The object considered here is given by the differential equation

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ -b \end{bmatrix} \mathrm{sgn}(x_2(t)) + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} u(t) \tag{53}$$

describing the mass $m > 0$, submitted to the bounded control $u(t) \in [-1, 1]$, with the discontinuous model of motion resistances $-b\mathrm{sgn}(x_2(t))$ with $0 < b < 1$. Then $x_1(t)$ and $x_2(t)$ refer to the position and the velocity of the mass, respectively.

The parameter b represents the whole array of factors of motion resistances, in particular friction with respect to surface and air, rolling, viscosity, etc., and due to the variety and complexity of each of these, one cannot speak of its 'actual' concrete value. In the case of underestimation, i.e. when $\hat{b} < b$, overregulations appear in the system, increasing the time for reaching the target as the difference $b - \hat{b}$ grows. On the other hand, for overestimation $b < \hat{b}$ the system generates much more effective sliding trajectories, which become less effective the greater the difference $\hat{b} - b$. For details, see the paper [25]. The procure of fixing a value for the estimator $\hat{b}$ should therefore take into account losses arising from the estimation error value and especially its sign, appropriate to the difference in effectiveness of the sliding trajectories and overregulations.

In turn, the mass m represents a concrete physical parameter, although for most practical tasks impossible to directly measure in a real-time regime. The error in estimating its value implies similar (but inverse) phenomena to the above described for the parameter b: underestimation $\hat{m} < m$ results in overregulation increasing the time for reaching the target more so as $m - \hat{m}$ increases, while overestimation $m < \hat{m}$ gives more effective sliding trajectories with greater efficiency as the difference $\hat{m} - m$ decreases. For details, see the paper [26]. Finally, the estimation of the parameter m should also account for the asymmetry of losses resulting in estimation error.

Joining the above considerations for the parameters b and m, it can be noted that these parameters represent completely different physical phenomena and are independent. The results of errors of their estimation are, however, correlated and lead to the same phenomena (overregulations and sliding trajectories) and in effect – as should be particularly underlined – they can cumulate or be partially eliminated.

Let us illustrate, therefore, the interpretation of particular coefficients of loss function (15). Let b be the first identified parameter, and m the second, i.e. $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b \\ m \end{bmatrix}$. In the case of the parameter b, overestimations are associated with more efficient sliding trajectories, and so the parameter a_l should be greater than a_r. The opposite holds true for the parameter m: the sliding trajectories occur through underestimation; therefore a_u ought to be greater than a_d. In the case of overestimation or underestimation of both parameters, the phenomena when the tendencies towards overregulations and sliding trajectories compensate each other are described by the parameters a_{ld} and a_{ru}, as opposed to the parameters a_{rd} and a_{lu}, which represent the sum of negative effects of cumulating overregulations and decreasing effectiveness of sliding trajectories.

A negative correlation was assumed between the values of the parameter b (characterizing among others friction) and wetness of a surface. The parameter m is not dependent on any other conditioning factors, which generally does not limit the usefulness of the method worked out, while due to the assumed (unconditional) independence of the parameters b and m assures conditional independence (see Appendix 2).

The simulation results show that the use of the method proposed here led to a reduction in time of reaching the target set of 3–10% with respect to the unconditional version using the mean value as an estimator, and 5–15% shorter than in the case of the classical (unconditional) sliding mode control,[27] where sliding trajectories had a notably larger number of switchings, which lowered their effectiveness.

Equally satisfying results have been obtained in initiated studies in medicine, for the case of hypotension, i.e. lowering cardiovascular pressure. Therapy consisting of taking one tablet of medicine containing two active ingredients appears to be significantly more effective than a traditional strategy of beginning treatment with monotherapy (prescribing one preparation) and next adding the second, then increasing the dosage of the latter until the desired cardiovascular pressure is obtained (see, e.g. the paper [28] also for rich subject literature). Furthermore, certain combinations of hypotensive preparations are more effective as far as prevention of organ complications is concerned, despite a similar performance in reducing pressure. The strength of the reaction to the medicine lowering pressure depends on many different personal and external factors [29] (at times – due to the current state of the patient – the nature of the drug's behaviour can even change). For this reason, it is also important to incorporate diverse conditioning factors, including current systolic and diastolic pressure as well as pulse, time of day and atmospheric pressure (continuous variables), as well as physical exertion and emotional state (categorical ordered variables). A fundamental problem with deciding on the therapy here is the definition of optimal dosages of preparations of hypotensive medicine combining two ingredients, on the basis of previously obtained measurements and current values for conditioning factors.

For both separately considered estimated quantities – i.e. of particular preparations of the two-ingredient medicine – potential overestimation is preferred to underestimation (which constitutes an asymmetric form of loss function with $a_r < a_l$ and $a_u < a_d$). The latter results in too small a dose of the medicine lowering pressure, and as a consequence, its insufficient reduction, in the extreme case a stroke. For the same reason, simultaneous underestimation of both these quantities is doubly dangerous, which implies a positive value of the parameter a_{ld}. Simultaneous overestimation, represented by the parameter a_{ru} is also unfavourable, although – thanks to the organism's adaptive mechanisms – it does not present such a large risk (i.e. $0 < a_{ru} < a_{ld}$). Overestimation of one preparation and underestimation of the second more or less compensate each other ($a_{rd} \cong 0 \cong a_{lu}$). Generally, the greater values of errors of estimation are adversely out of proportion, which justifies the square form of the performance index, with the assumptions concerning the parameters $a_{\bullet}$ and $a_{\bullet\bullet}$ formulated below formula (15). Thanks to the fixing of the estimators for both ingredients – conditioned by the form of the loss function defined on the basis of the above circumstances and precised by current values for conditioning variables – it is possible to appropriately individualize amounts of the preparations for the two-ingredient hypotensive medicine, and especially suitably fit them to specific, actual, personal situations. The value of the resulting estimator is moved with respect to the mean of obtained measurements, in the direction associated with the least losses, with proper influence of current values of conditioning factors.

8. Summary

This paper presents the algorithm for calculating the conditioning value of a parameters' vector, 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 investigated algorithm – together with the subject procedures from the quoted literature – is ready for direct use without any additional laborious research or calculations. The presented concept is universal in nature and can be applied in a wide range of tasks in science, engineering, economy and management, environmental and social issues, biomedicine and other related fields. The results have been verified positively based on generated and real data for practical problems from control engineering and medicine areas.

And finally, on a personal note. In a previous paper [8], we wrote that in the multidimensional case 'both the analytical criteria for optimal parameter values as well as their later numerical implementation become too complicated for practical application, given today's possibilities'. Fortunately, upon introducing the assumption concerning the conditional independence of identified parameters, while maintaining the correlation of losses resulting from estimation errors, the above statement turned out to be too pessimistic. Moreover, the simulations carried out show the proposed algorithm to be fairly robust to that assumption and that in practice it can be successfully applied, even if not absolutely fulfilled.

Acknowledgements

Our heartfelt thanks go to our colleague Dr Aleksander Mazgaj, with whom we commenced the research presented here. With his consent, this text also contains results of joint research. We are also very grateful to Professor Antoni Leon Dawidowicz for his valuable advice and inspiration.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Walter E, Pronzato I. Identification of parametric models. Berlin: Springer-Verlag; 1997.
[2] Lehmann EL. Theory of point estimation. New York: Wiley; 1983.
[3] Kulczycki P, Hryniewicz O, Kacprzyk J, eds. Techniki informacyjne w badaniach systemowych. Warsaw: WNT; 2007.
[4] Berger JO. Statistical decision theory. New York: Springer-Verlag; 1980.
[5] Zellner A. Bayesian estimation and prediction using asymmetric loss function. J Am Stat Assoc. 1986;81:446–451.
[6] McCullough BD. Optimal prediction with a general loss function. J Comb Inf Syst Sci. 2000;25:207–221.
[7] Kulczycki P, Mazgaj A. Parameter identification for asymmetrical polynomial loss function. Inf Technol Control. 2009;38:51–60 [Errata: vol. 38:167–168].
[8] Kulczycki P, Charytanowicz M. Conditional parameter identification with different losses of under- and overestimation. Appl Math Model. 2013;37:2166–2177.
[9] Kulczycki P, Charytanowicz M. Warunkowa wielowymiarowa identyfikacja wartosci parametrow przy niesymetrycznych i skorelowanych stratach bledow estymacji. XVIII Krajowa Konferencja Automatyki; 2014 September 8–10; Wroclaw; CD: 62.
[10] Kulczycki P, Charytanowicz M. Conditional multidimensional parameter identification with asymmetric correlated losses of estimation errors. The 21st international conference on neural information processing; 2014 November 3–6; Kuching (Malaysia).Loo CK, Yap KS, Wong KW, Teoh A, Huang K, editors. Neural information processing. Lecture notes in computer science, vol. II. Berlin: Springer. p. 287–294.
[11] Kulczycki P. Estymatory jadrowe w analizie systemowej. Warsaw: WNT; 2005.
[12] Wand MP, Jones MC. Kernel smoothing. London: Chapman and Hall; 1995.
[13] Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.
[14] Simonoff JS. Smoothing methods in statistics. New York: Springer; 1996.
[15] Stoer J, Bulirsch R. Introduction to numerical analysis. New York: Springer-Verlag; 2002.
[16] Ahmad IA, Cerrito PB. Nonparametric estimation of joint discrete-continuous probability densities with applications. J Stat PlanInference. 1994;41:349–364.
[17] Wang M, van Ryzin J. A class of smooth estimators for discrete distributions. Biometrica. 1981;68:301–309.
[18] Aitken CGG. Kernel methods for the estimation of discrete distributions. J Stat Comput Simul. 1983;16:189–200.
[19] Gaosheng J, Rui L, Zhongwen L. Nonparametric estimation of multivariate CDF with categorical and continuous data. Adv Econ. 2009;25:291–318.
[20] Li Q, Racine JS. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. J Bus Econ Stat. 2008;26:423–434.
[21] Ouyang D, Li Q, Racine JS. Cross-validation and the estimation of probability distributions with categorical data. J Nonparametr Stat. 2006;18:69–100.
[22] Deuflhard P. Newton methods for nonlinear problems. Affine invariance and adaptive algorithms. Berlin: Springer; 2004.
[23] Kelley CT. Solving nonlinear equations with Newton's method. Philadelphia, PA: SIAM. 2003.
[24] Sontag EG. Mathematical control theory. New York: Springer; 1998.
[25] Kulczycki P. Fuzzy controller for mechanical systems. IEEE Trans Fuzzy Syst. 2000;8:645–652.
[26] Kulczycki P, Wisniewski R. An algorithm for Bayes parameter identification. J Dyn Syst Meas Control. 2002;123:611–614.
[27] Utkin VI. Sliding modes in control optimization. Berlin: Springer-Verlag; 1992.
[28] Gradman AH, Basile JN, Carter BL, Bakris GL. Combination therapy in hypertension. J Am Soc Hypertens. 2010;4:90–98.
[29] Frank J. Managing hypertension using combination therapy. Am Family Physician. 2008;77:1279–1286.
[30] Dawid AP. Conditional independence in statistical theory. J R Stat Soc Ser B. 1979;41:1–31.

Appendix 1 (sufficient condition for optimization)

As mentioned in Section 3, this appendix will show that the solution of Equations (22) and (23) exists, is unique and is a global minimum of the function l_B.

The form of the function l_B given by formula (19) implies that the derivatives $\partial l_\mathrm{B}/\partial \hat{y}_1$ and $\partial l_\mathrm{B}/\partial \hat{y}_2$ are continuous, while thanks to assumptions (17) and (18), based on the equalities (20) and (21) we have

$$\lim_{\substack{\hat{y}_1 \to -\infty \\ \hat{y}_2 \to -\infty}} \frac{\partial l_\mathrm{B}(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_1} < 0 \quad \text{and} \quad \lim_{\substack{\hat{y}_1 \to \infty \\ \hat{y}_2 \to \infty}} \frac{\partial l_\mathrm{B}(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_1} > 0, \tag{A1}$$

$$\lim_{\substack{\hat{y}_1\to-\infty\\ \hat{y}_2\to-\infty}} \frac{\partial l_{\mathrm{B}}(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2} < 0 \quad \text{and} \quad \lim_{\substack{\hat{y}_1\to\infty\\ \hat{y}_2\to\infty}} \frac{\partial l_{\mathrm{B}}(\hat{y}_1,\hat{y}_2)}{\partial \hat{y}_2} > 0, \tag{A2}$$

therefore Equations (22) and (23) are fulfilled by at least one element $\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}$. It is denoted by $\begin{bmatrix}\hat{y}_1^*\\ \hat{y}_2^*\end{bmatrix}$.

In order to establish whether the solution of Equations (22) and (23) is unique and is a global minimum of the function l_{B}, the sign of partial derivatives will now be examined. Due to the symmetry of Equations (22) and (23), analysis will be carried out for the former – in the case of the latter considerations are analogous.

When introducing the number $\Delta \in \mathbb{R}$, for any $\hat{y}_1 \in \mathbb{R}$ one can express $\hat{y}_1 = \hat{y}_1^* + \Delta$. Let $\hat{y}_2$ be freely fixed. Equation (22) may then be equivalently denoted as

$$\begin{aligned}
\frac{\partial l_B}{\partial \hat{y}_1}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) = {} & \left(a_{ld}\int_{\hat{y}_1^*+\Delta}^{\infty} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + a_{rd}\int_{-\infty}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\\
& \times \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + \left(a_{lu}\int_{\hat{y}_1^*+\Delta}^{\infty} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + a_{ru}\int_{-\infty}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\\
& \times \int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + 2a_l\int_{\hat{y}_1^*+\Delta}^{\infty} (\hat{y}_1^* + \Delta - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\\
& + 2a_r\int_{-\infty}^{\hat{y}_1^*+\Delta} (\hat{y}_1^* + \Delta - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1.
\end{aligned} \tag{A3}$$

Let first $\Delta > 0$. By applying additivity of the integral with respect to the integration set, the above formula yields

$$\begin{aligned}
\frac{\partial l_B}{\partial \hat{y}_1}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) = {} & \left(a_{ld}\int_{\hat{y}_1^*}^{\infty} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 - a_{ld}\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right.\\
& \left. + a_{rd}\int_{-\infty}^{\hat{y}_1^*} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + a_{rd}\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\\
& \times \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + \left(a_{lg}\int_{\hat{y}_1^*}^{\infty} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right.\\
& - a_{lu}\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + a_{pu}\int_{-\infty}^{\hat{y}_1^*} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\\
& \left. + a_{ru}\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2\\
& + 2a_l\int_{\hat{y}_1^*}^{\infty} (\hat{y}_1^* - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + 2a_l\Delta\int_{\hat{y}_1^*}^{\infty} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\\
& - 2a_l\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} (\hat{y}_1^* - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 - 2a_l\Delta\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\\
& + 2a_r\int_{-\infty}^{\hat{y}_1^*} (\hat{y}_1^* - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + 2a_r\Delta\int_{-\infty}^{\hat{y}_1^*} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\\
& + 2a_r\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} (\hat{y}_1^* - y_1) f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + 2a_r\Delta\int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1.
\end{aligned} \tag{A4}$$

Because the value $\hat{y}_1^*$ is assumed to fulfil Equation (22), this formula is simplified to the form:

$$\begin{aligned}
\frac{\partial l_{\mathrm{B}}}{\partial \hat{y}_1}\left(\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix}\right) = {} & \left(-a_{ld}\int_{\hat{y}_1}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1 + a_{rd}\int_{\hat{y}_1}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\\
& \times \int_{\hat{y}_2}^{\infty} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2 + \left(a_{ru}\int_{\hat{x}_1}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right.\\
& \left. - a_{lu}\int_{\hat{y}_1}^{\hat{y}_1^*+\Delta} f_{Y_1|W=w^*}(y_1)\,\mathrm{d}y_1\right)\int_{-\infty}^{\hat{y}_2} (\hat{y}_2 - y_2) f_{Y_2|W=w^*}(y_2)\,\mathrm{d}y_2
\end{aligned}$$

$$+ 2a_l\Delta \int_{\hat{y}_1^*+\Delta}^{\infty} f_{Y_1|W=w^*}(y_1)\ \mathrm{d}y_1 - 2a_l \int_{\hat{y}_1}^{\hat{y}_1^*+\Delta} (\hat{y}_1^* - y_1) f_{Y_1|W=w^*}(y_1)\ \mathrm{d}y_1$$

$$+ 2a_r\Delta \int_{-\infty}^{\hat{y}_1^*} f_{Y_1|W=w^*}(y_1)\ \mathrm{d}y_1 + 2a_r \int_{\hat{y}_1^*}^{\hat{y}_1^*+\Delta} (\hat{y}_1^* + \Delta - y_1) f_{Y_1|W=w^*}(y_1)\ \mathrm{d}y_1. \tag{A5}$$

Due to the assumptions concerning coefficients of loss function (15), the expression inside the first parentheses is negative. Its multiplication by the subsequent integral, whose value is also negative, implies a resulting positive value. A similar analysis of the remaining expressions shows that the whole expression (A5) is positive for any $\Delta > 0$. In the same way, one can prove that for any $\Delta < 0$, this expression is negative.

Identical treatment can be carried out for Equation (23). This is final proof that the solution for conditions (22)–(23) exists and is unique with the Bayes loss function l_B assuming here its global minimum.

Appendix 2 (comments on conditional independence)

In this section, the property of conditional independence will be briefly commented, as it goes slightly beyond intuition, by too impetuously transferring the ubiquitous classical unconditional version. Consider two random variables Y_1 and Y_2; their (unconditional) independence means that knowledge of the realization of one of them does not provide any information concerning the value of the realization of the other. If however both are associated with the same conditioning random variable W, then additionally the knowledge of the realization value of one of the variables Y_1 or Y_2 allows inference of the realization value of the conditioning variable, and thereby of the realization value of the other. Perhaps the information obtained thus will not contribute much, but it will be something. Intuitively, the notion of an unconditional independence transfers to a conditional approach if:

(a) the variables Y_1 and Y_2 are dependent (unconditionally) in such a way that their direct mutual influence is compensated by information carried by the conditioning variable W;
(b) one of the variables Y_1 and Y_2 is independent of W.

Let us illustrate the above, assuming that the composition $X = \begin{bmatrix} Y_1 \\ Y_2 \\ W \end{bmatrix}$ is of Gauss distribution with the expectation and covariance matrixes

$$E_X = \begin{bmatrix} E_{Y_1} \\ E_{Y_2} \\ E_W \end{bmatrix}, \quad \mathrm{COV}_X = \begin{bmatrix} \mathrm{var}_{Y_1} & \mathrm{cov}_{Y_1Y_2} & \mathrm{cov}_{Y_1W} \\ \mathrm{cov}_{Y_2Y_1} & \mathrm{var}_{Y_2} & \mathrm{cov}_{Y_2W} \\ \mathrm{cov}_{WY_1} & \mathrm{cov}_{WY_2} & \mathrm{var}_W \end{bmatrix}. \tag{A6}$$

Density is therefore given as

$$f_X(x) = f\left(\begin{bmatrix} y_1 \\ y_2 \\ w \end{bmatrix}\right) \frac{1}{(2\pi)^{3/2}\sqrt{\det(\mathrm{COV}_X)}} \mathrm{e}^{(((x-E_X)^\mathrm{T}\mathrm{COV}_X^{-1}(x-E_X))/2)}. \tag{A7}$$

The variables Y_1 and Y_2 become conditionally independent with respect to W, when the function f can be expressed as the product of the functions with the arguments y_1 and w, as well as y_2 and w; so if the expression in the exponent of formula (A7) will not contain the products y_1y_2. Therefore, in the matrix $A = \mathrm{COV}_X^{-1}$, the elements a_{12} and a_{21} should be equal to zero, i.e.

$$a_{12} = a_{21} = \frac{1}{\det(\mathrm{COV}_X)} \begin{vmatrix} \mathrm{cov}_{Y_1Y_2} & \mathrm{cov}_{Y_1W} \\ \mathrm{cov}_{WY_2} & \mathrm{var}_W \end{vmatrix} = 0; \tag{A8}$$

and finally taking into account the symmetry of the matrix COV_X

$$\mathrm{cov}_{Y_1Y_2}\ \mathrm{var}_W = \mathrm{cov}_{Y_1W}\ \mathrm{cov}_{Y_2W}. \tag{A9}$$

The above condition fulfils the covariance matrixes defined in formulas (44)–(48) and exemplifies the case (a) formulated above. Meanwhile in system (52), independence from the conditioning factor – the mass m – results in the case (b), presented earlier, appearing in this example.

Further, detailed information regarding conditional independence can be found in the survey paper.[30]