Applied Mathematical Modelling 37 (2013) 2166–2177

Contents lists available at SciVerse ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Conditional parameter identification with different losses of under- and overestimation

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

[a] Polish Academy of Sciences, Systems Research Institute, Centre of Information Technology for Data Analysis Methods, Poland
[b] Cracow University of Technology, Department of Automatic Control and Information Technology, Poland
[c] Catholic University of Lublin, Institute of Mathematics and Computer Science, Poland

ARTICLE INFO

Article history:
Received 29 August 2011
Received in revised form 27 April 2012
Accepted 7 May 2012
Available online 22 May 2012

Keywords:
Parameter identification
Asymmetrical losses of estimation errors
Conditional factors
Nonparametric estimation
Statistical kernel estimators
Numerical algorithm

ABSTRACT

In many scientific and practical tasks, the classical concepts for parameter identification are satisfactory and generally applied with success, although many specialized problems necessitate the use of methods created with specifically defined assumptions and conditions. This paper investigates the method of parameter identification for the case where losses resulting from estimation errors can be described in polynomial form with additional asymmetry representing different results of under- and overestimation. Most importantly, the method presented here considers the conditionality of this parameter, which in practice means its significant dependence on other quantities whose values can be obtained metrologically. To solve a problem in this form the Bayes approach was used, allowing a minimum expected value of losses to be achieved. The methodology was based on the nonparametric technique of statistical kernel estimators, which freed the investigated procedure from forms of probability distributions characterizing both the parameter under investigation and conditioning quantities. As a result an algorithm is presented, ready for direct use without further intensive research and calculations.

1. Introduction

Parameter identification [1], i.e. assigning a concrete value to a parameter present in a model, despite its very traditional nature, has still great significance in modern scientific and applicational problems. Moreover, its importance continuously increases together with the dominance of model-based methods and the growing, often specific, demands made on models used in science and practice. At the same time, the increasing complexity and novelty of current methods is accompanied by a decrease in the classical understanding of parameter identification as a task of fixing a concrete value of a parameter which exists objectively in reality but is unknown. Here, through investigating, the researcher attempts to get as close as possible to this "true" value. In fact more frequently in contemporary models, their particular parameters describe an entire range of complex phenomena, simplified in a model to one parameter, existing only formally – without concrete physical form. In this situation the quality of parameter identification cannot be evaluated by classical means, obtaining a value as near as possible to an imagined "true" parameter value (since it does not exist), but rather by accounting for the influence of particular parameter values on a considered system, whose part is the investigated model. This moves the mathematical apparatus applied here – present within point estimation – from classical mathematical statistics [2], towards the currently intensively-studied data analysis [3]. Fortunately, the development of modern sophisticated and often specific methods of

* Corresponding author at: Polish Academy of Sciences, Systems Research Institute, Centre of Information Technology for Data Analysis Methods, Poland.
E-mail addresses: kulczycki@ibspan.waw.pl, kulczycki@pk.edu.pl (P. Kulczycki).

0307-904X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.apm.2012.05.007

parameter identification is facilitated by the dynamic expansion of contemporary computer technology, supported on the theoretical side by the procedures of advanced information technology [4].

The subject of this paper is an algorithm for parameter identification, i.e. estimation of the value of a parameter occurring in a model, based on four premises:

1. minimization of expected value of losses arising from estimation errors, unavoidable in practice;
2. asymmetry of those losses, i.e. allowing for situations where losses occurring through underestimation are substantially different from losses resulting from overestimation;
3. arbitrariness of probability distributions appearing in the problem;
4. and finally–worth particularly highlighting–conditionality of an identified parameter, that is its significant dependence on a factor (or factors), with values that can be in practice obtained metrologically.

The realization of the first will be through application of the Bayes approach [5].

The second by assuming the loss function resulting from estimation errors, in the asymmetrical form

$$l(\hat{y}, y) = \begin{cases} (-1)^k a_l (\hat{y} - y)^k & \text{for} \quad \hat{y} - y \leqslant 0, \\ a_r (\hat{y} - y)^k & \text{for} \quad \hat{y} - y \geqslant 0, \end{cases} \tag{1}$$

with the given degree $k \in \mathbf{N} \setminus \{0\}$, where the coefficients a_l and a_r are positive, while y and $\hat{y}$ denote the values of the parameter under consideration and its estimator, respectively. The fact that the coefficients a_l and a_r may differ causes an asymmetry of the above function and enables the inclusion of different losses implied by over- and underestimation of the examined parameter. Limiting the form of function (1) to a polynomial seems not to decrease the generality of considerations in practical applications, offering an effective compromise between precision and complexity of results obtained. Moreover the possibility of change of the polynomial degree k – with respect to that resulting from fundamental research – allows a differing scale of protection against large estimation errors.

The third aspect is realized by applying nonparametric methodology of statistical kernel estimators [6–8] for calculating probability characteristics.

Lastly – and worth highlighting once more – this paper is aimed at the conditional approach, i.e. where the value of the estimated parameter is strongly dependent on a conditional factor, for example in engineering practice it is often a current temperature. If the value of such a factor is metrologically available, then its inclusion can make the used model significantly more precise.

The goal of this paper is the provision of an algorithm for calculating a conditional parameter value, optimal in the sense of minimum expectation value of losses, in particular those different for under- and overestimation. The above value is determined for a fixed (most often current) value of a conditional factor, based on measurements of this parameter obtained earlier for different conditioning values. The algorithm is comprehensive and can be applied directly without detailed knowledge of theoretical aspects, laborious research or analytical calculations. It is sufficient data to take only the measurements of pairs of the model parameter value, and the conditional factor value for which this parameter value was obtained, as well as the quantities introduced in formula (1): the degree k and the ratio of coefficients a_l/a_r.

Thus, Section 2 outlines the statistical kernel estimators method. The algorithm worked out is described in Sections 3 and 4, with the asymmetrical linear case in Section 3.1, the asymmetrical quadratic in Section 3.2, and the asymmetrical polynomial (in particular cubic) in Section 3.3. Finally, Section 5 presents the results of experimental verification of the investigated procedure. Section 6 provides a summary of the presented method.

The preliminary version of this paper was presented as [9].

2. Preliminaries: statistical kernel estimators

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

$$x_1, x_2, \ldots, x_m, \tag{2}$$

in its basic form is defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

where $m \in \mathbf{N} \setminus \{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K : \mathbf{R}^n \to [0, \infty)$ of unit integral $\int_{\mathbf{R}^n} K(x)\, dx = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. The interpretation of the above definition is illustrated in Fig. 1 for a one-dimensional random variable. In the case of the single realization x_i, the function K (transposed along the vector x_i and scaled by the coefficient h) represents the approximation of distribution of the random variable X having obtained the value x_i. For m independent realiza-

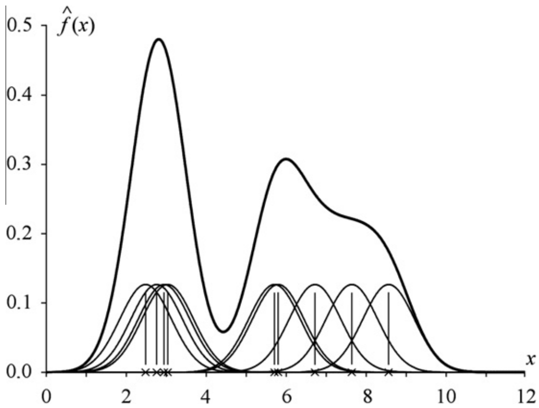

Fig. 1. Kernel estimator (3) for one-dimensional random variable ($n = 1$) and 9-element sample ($m = 9$).

tions $x_1, x_2, \ldots, x_m$, this approximation takes the form of a sum of these single approximations. The constant $1/mh^n$ enables the condition $\int_{\mathbf{R}^n} \hat{f}(x)dx = 1$, required of the density of a probability distribution.

The choice of a form of the kernel K and the calculation of the smoothing parameter h is made most often with the criterion of the mean integrated square error.

The results obtained with this criterion point to the form of the kernel having – from a statistical point of view – no practical meaning. Thanks to this, when assuming the function K it becomes possible to take into account primarily properties of the estimator $\hat{f}$ obtained (e.g. its continuity or differentiability) or calculational aspects (e.g. the possibility of analytically calculating an integral), advantageous from the point of view of the applicational problem under investigation. For broader discussion see [6] – Section 3.1.3, [8] – 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 in analysis – is often favored in practical problems. The n-dimensional product kernel K can be expressed as

$$K(x) = K\left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}\right) = \mathbf{K}_1(\mathbf{x}_1)\mathbf{K}_2(\mathbf{x}_2)\cdots\mathbf{K}_n(\mathbf{x}_n), \tag{4}$$

where $\mathbf{K}_i$ for $i = 1, 2, \ldots, n$ denotes the previously-mentioned one-dimensional kernels, while the expression h^n appearing in the basic formula (3) should be replaced by $h_1 \cdot h_2 \cdot \ldots \cdot h_n$, the product of the smoothing parameters for particular coordinates.

The fixing of the smoothing parameter h has significant meaning for quality of estimation. Fortunately – from the applicational point of view – many suitable procedures for calculating the value of the parameter h on the basis of random sample (2) have been worked out. For broader discussion of the above tasks see [6–8]. In particular, for the one-dimensional case, the simple and effective plug-in method ([6] – Section 3.1.5, [8] – Section 3.6.1) is especially recommended. Of course this method can also be applied in the n-dimensional case when product kernel (4) is used, separately for each coordinate.

Practical applications may also use additional procedures generally improving the quality of estimator (3). For the method presented in this paper, the modification of the smoothing parameter ([6] – Section 3.1.6, [7] – Section 5.3.1) is strongly 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 dimension $n_Y + n_W$. Assume that distributions of the variables X and, in consequence, W have densities, denoted below as $f_X : \mathbf{R}^{n_Y+n_W} \to [0,\infty)$ and $f_W : \mathbf{R}^{n_W} \to [0,\infty)$, respectively. Let also be given the so-called conditioning value, that is the fixed value of conditioning random variable $w^* \in \mathbf{R}^{n_W}$, such that

$$f_W(w^*) > 0. \tag{5}$$

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

$$f_{Y|W=w^*}(y) = \frac{f_X(y, w^*)}{f_W(w^*)} \text{ for every } y \in \mathbf{R}^{n_Y} \tag{6}$$

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

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \ldots, \begin{bmatrix} y_m \\ w_m \end{bmatrix}, \tag{7}$$

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. Using the methodology presented in the first part of the section below, on the basis of sample (7) one can calculate $\hat{f}_X$, i.e. the kernel estimator of density of the random variable X probability distribution, while the sample

$$w_1, w_2, \ldots, w_m \tag{8}$$

gives $\hat{f}_W$ – the kernel density estimator for the conditioning variable W. The kernel estimator of conditional density of the random variable Y probability distribution for the conditioning value w^*, is defined then – a natural consequence of formula (6) – as the function $\hat{f}_{Y|W=w^*} : \mathbf{R}^{n_Y} \to [0, \infty)$ given by

$$\hat{f}_{Y|W=w^*}(y) = \frac{\hat{f}_X(y, w^*)}{\hat{f}_W(w^*)}. \tag{9}$$

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 (5) is fulfilled for any $w^* \in \mathbf{R}^{n_W}$.

In the case when for the estimators $\hat{f}_X$ and $\hat{f}_W$ the product kernel (4) 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 (9) can then be specified to the form

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{n_Y} \end{bmatrix}\right) = \frac{\frac{1}{h_1 h_2 \ldots h_{n_Y}} \sum_{i=1}^m \mathbf{K}_1\left(\frac{\mathbf{y}_1 - y_{i,1}}{h_1}\right) \mathbf{K}_2\left(\frac{\mathbf{y}_2 - y_{i,2}}{h_2}\right) \ldots \mathbf{K}_{n_Y}\left(\frac{\mathbf{y}_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right) \mathbf{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathbf{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathbf{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right)}{\sum_{i=1}^m \mathbf{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathbf{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathbf{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right)}, \tag{10}$$

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

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W} \end{bmatrix} \quad \text{and} \quad y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_Y} \end{bmatrix}, \; w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{11}$$

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

$$d_i = \mathbf{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathbf{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathbf{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right). \tag{12}$$

Thanks to the assumption of positive values for the kernels $\mathbf{K}_{n_Y+1}, \mathbf{K}_{n_Y+2}, \ldots, \mathbf{K}_{n_Y+n_W}$, these parameters are also positive. So the kernel estimator of conditional density (10) can be presented in the form

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{n_Y} \end{bmatrix}\right) = \frac{1}{h_1 h_2 \ldots h_{n_Y} \sum_{i=1}^m d_i} \sum_{i=1}^m d_i \mathbf{K}_1\left(\frac{\mathbf{y}_1 - y_{i,1}}{h_1}\right) \mathbf{K}_2\left(\frac{\mathbf{y}_2 - y_{i,2}}{h_2}\right) \ldots \mathbf{K}_{n_Y}\left(\frac{\mathbf{y}_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right). \tag{13}$$

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 (13) 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^*. The factor $\sum_{i=1}^m d_i$ norms the value of the estimator with the aim of ensuring a unit integral, i.e. the condition $\int_{\mathbf{R}^{n_Y}} \hat{f}_{Y|W=w^*}(y)\, dy = 1$ for any $w^* \in \mathbf{R}^{n_W}$.

Returning to the subject of this article, described in the Introduction, in the case of estimation of a single parameter, the random variable Y is one-dimensional ($n_Y = 1$). This will be investigated further in the presented paper. However when one estimates a number of conditionally correlated parameters, then n_Y becomes equal to their number – this case will be commented upon at the end of Section 6.

More details concerning kernel estimators can be found in the books [6–8]. See also [10] to find classic bibliography. Exemplary applications are presented in the publications [11–17].

3. Main results

3.1. Linear case

Let the parameter under investigation, whose value is to be estimated, denoted by $y \in \mathbf{R}$, be treated as the value of the 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 parameter y, i.e. $y_1, y_2, \ldots, y_m$, obtained for the values $w_1, w_2, \ldots, w_m$ of the conditional variable, respectively. Finally, let $w^* \in \mathbf{R}^{n_W}$ denote any fixed conditioning value. The goal is to calculate the estimator of this parameter, denoted by $\hat{y}_{w^*}$, optimal in the sense of minimum expected value of losses arising from errors of estimation, for conditioning value w^*. The case considered in this subsection is such that loss function (1) can be specified to the following asymmetrical linear form:

$$l(\hat{y}_{w^*}, y) = \begin{cases} -a_l\,(\hat{y}_{w^*} - y) & \text{for} \quad \hat{y}_{w^*} - y \leqslant 0, \\ a_r\,(\hat{y}_{w^*} - y) & \text{for} \quad \hat{y}_{w^*} - y \geqslant 0, \end{cases} \tag{14}$$

while the coefficients a_l and a_r are positive and not necessarily equal to each other.

In order to solve such a task, the Bayes decision rule will be used [5]. The minimum expected value of losses arising from estimation errors occurs when the value is a solution of the following equation with the argument $\hat{y}_{w^*}$:

$$\int_{-\infty}^{\hat{y}_{w^*}} f_{Y|W=w^*}(y)\,dy - \frac{a_l}{a_l + a_r} = 0, \tag{15}$$

where $f_{Y|W=w^*}$ denotes the density of distribution of the random variable Y representing the uncertainty of the parameter in question, for conditioning value w^*. Since $0 < a_l/(a_l + a_r) < 1$, a solution for the above equation exists, and if the function $f_{Y|W=w^*}$ has connected support, this solution is unique. Moreover, thanks to equality $\frac{a_l}{a_l+a_r} = \frac{a_l/a_r}{a_l/a_r+1}$, it is not necessary to identify the parameters a_l and a_r separately, rather only their ratio.

The identification of the density $f_{Y|W=w^*}$ will be carried out using statistical kernel estimators, presented in Section 2, with the – convenient here – form (13). Then as $\mathbf{K}_1$ (note that $n_Y = 1$) one should choose a continuous kernel of positive values, and also so that the function $\mathbf{I}:\mathbf{R} \to \mathbf{R}$ such that $\mathbf{I}(x) = \int_{-\infty}^{x} \mathbf{K}_1(y)\,dy$ can be expressed by a relatively simple analytical formula. In consequence, this results in a similar property regarding the function $\mathbf{U}_i : \mathbf{R} \to \mathbf{R}$ for any fixed $i = 1, 2, \ldots, m$ defined as

$$\mathbf{U}_i(\hat{y}_{w^*}) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_{w^*}} \mathbf{K}_1\left(\frac{y - x_i}{h_1}\right) dy. \tag{16}$$

Criterion (15) can be expressed then equivalently in the form of

$$\sum_{i=1}^{m} d_i \mathbf{U}_i(\hat{y}_{w^*}) - \frac{a_l}{(a_l + a_r)} \sum_{i=1}^{m} d_i = 0. \tag{17}$$

If the left side of the above equation is denoted by $L(\hat{y}_{w^*})$, then $\lim_{\hat{y}_{w^*} \to -\infty} L(\hat{y}_{w^*}) < 0$, $\lim_{\hat{y}_{w^*} \to \infty} L(\hat{y}_{w^*}) > 0$, the function L is (strictly) increasing and its derivative may be simply expressed by

$$L'(\hat{y}_{w^*}) = \sum_{i=1}^{m} d_i \mathbf{K}_1\left(\frac{\hat{y}_{w^*} - x_i}{h_1}\right). \tag{18}$$

In this situation, the solution of criterion (15) can be effectively calculated on the basis of Newton's algorithm [18] as the limit of the sequence $\{\hat{y}_{w^*,j}\}_{j=0}^{\infty}$ defined by

$$\hat{y}_{w^*,0} = \frac{\sum_{i=1}^{m} d_i y_i}{\sum_{i=1}^{m} d_i}, \tag{19}$$

$$\hat{y}_{w^*,j+1} = \hat{y}_{w^*,j} - \frac{L(\hat{y}_{w^*,j})}{L'(\hat{y}_{w^*,j})} \quad \text{for} \quad j = 0, 1, \ldots, \tag{20}$$

with the functions L and L' being given by dependencies (17) and (18), whereas a stop criterion takes on the form

$$|\hat{y}_{w^*,j} - \hat{y}_{w^*,j-1}| \leqslant 0.01\,\hat{\sigma}_Y, \tag{21}$$

while $\hat{\sigma}_Y$ denotes the estimator of the standard deviation of the random variable Y.

3.2. Quadratic case

The conditionings of the problem investigated in this subsection are similar to the previous one, although asymmetric linear form of the loss function (14) is substituted by the asymmetric quadratic:

$$l(\hat{y}_{w^*}, y) = \begin{cases} a_l\,(\hat{y}_{w^*} - y)^2 & \text{for} \quad \hat{y}_{w^*} - y \leqslant 0, \\ a_r\,(\hat{y}_{w^*} - y)^2 & \text{for} \quad \hat{y}_{w^*} - y \geqslant 0, \end{cases} \tag{22}$$

while the coefficients a_l and a_r are positive and not necessarily equal to each other. The minimum expected value of losses arising from estimation errors can in this case be calculated for the value $\hat{y}_{w^*}$ being a solution of the equation

$$(a_l - a_r)\int_{-\infty}^{\hat{y}_{w^*}} (\hat{y}_{w^*} - y) f_{Y|W=w^*}(y)\, dy - a_l \int_{-\infty}^{\infty} (\hat{y}_{w^*} - y) f_{Y|W=w^*}(y)\, dy = 0. \tag{23}$$

This solution exists and is unique. As in the linear case, dividing the above equation by a_r, note that it is necessary to identify only the ratio of the parameters a_l and a_r.

Using kernel estimators in form (13) to identify the density $f_{Y|W=w^*}$, one can design an effective numerical algorithm to this end. Let, therefore, a continuous kernel $\mathbf{K}_1$ of positive values, fulfilling the condition

$$\int_{-\infty}^{\infty} y\,\mathbf{K}_1(y)\, dy < \infty \tag{24}$$

be given. Besides the functions $\mathbf{U}_i$ introduced by the dependence (16), let for any fixed $i = 1, 2, \ldots, m$ the functions $\mathbf{V}_i : \mathbf{R} \to \mathbf{R}$ be defined as

$$\mathbf{V}_i(\hat{y}_{w^*}) = \frac{1}{h_1}\int_{-\infty}^{\hat{y}_{w^*}} y\mathbf{K}_1\left(\frac{y - y_i}{h_1}\right) dy. \tag{25}$$

The kernel $\mathbf{K}_1$ should be chosen so that – apart from the requirements formulated above – the function $\mathbf{J} : \mathbf{R} \to \mathbf{R}$ such that $\mathbf{J}(x) = \int_{-\infty}^{x} y\,\mathbf{K}_1(y)\, dy$ be expressed by a convenient analytical formula.

Criterion (23) can then be described equivalently as

$$\sum_{i=1}^{m} d_i[(a_l - a_r)(\hat{y}_{w^*}\,\mathbf{U}_i(\hat{y}_{w^*}) - \mathbf{V}_i(\hat{y}_{w^*})) + a_l y_i] - m a_l \hat{y}_{w^*} \sum_{i=1}^{m} d_i = 0. \tag{26}$$

If the left side of the above formula is denoted by $L(\hat{y}_{w^*})$, then – using the equality $\mathbf{V}_i'(\hat{y}_{w^*}) = \hat{y}_{w^*}\,\mathbf{U}_i'(\hat{y}_{w^*})$ directly resulting from dependencies (16) and (25) – one can express the value of its derivative as

$$L'(\hat{y}_{w^*}) = \sum_{i=1}^{m} d_i[(a_l - a_r)\,\mathbf{U}_i(\hat{y}_{w^*})] - m a_l \sum_{i=1}^{m} d_i. \tag{27}$$

In this situation, the solution of criterion (23) can be calculated numerically on the basis of Newton's algorithm (19)–(21) with the functions L and L' defined by dependencies (26) and (27).

3.3. Higher degree polynomial case

In this subsection, the linear and quadratic approaches presented earlier will be supplemented with the polynomial case, that is where the loss function is an asymmetrical monomial of the order $k \geqslant 3$, i.e.

$$l(\hat{y}_{w^*}, y) = \begin{cases} (-1)^k a_l\,(\hat{y}_{w^*} - y)^k & \text{for} \quad \hat{y}_{w^*} - y \leqslant 0, \\ a_r\,(\hat{y}_{w^*} - y)^k & \text{for} \quad \hat{y}_{w^*} - y \geqslant 0, \end{cases} \tag{28}$$

while the coefficients a_l and a_r are positive, and may differ. Criterion for the optimal estimator takes on here the form of the following equation with the argument $\hat{y}_{w^*}$:

$$(-1)^k a_l \int_{\hat{y}_{w^*}}^{\infty} (\hat{y}_{w^*} - y)^{k-1} f_{Y|W=w^*}(y)\, dy + a_r \int_{-\infty}^{\hat{y}_{w^*}} (\hat{y}_{w^*} - y)^{k-1} f_{Y|W=w^*}(y)\, dy = 0. \tag{29}$$

The solution of the above equation exists and is unique. When the statistical kernel estimators are used with respect to the density $f_{Y|W=w^*}$, it is possible again to create an efficient numerical algorithm enabling equation (29) to be solved. The kernel $\mathbf{K}_1$ should here be continuous, of positive values and fulfilling the following condition:

$$\int_{-\infty}^{\infty} y^{k-1}\mathbf{K}_1(y)\, dy < \infty. \tag{30}$$

For clarity of presentation, the cubic case $k = 3$ is presented below in detail. Thus, formula (29) is equivalent to

$$(a_l + a_r)\left(\hat{y}_{w^*}^2 \int_{-\infty}^{\hat{y}_{w^*}} f_{Y|W=w^*}(y)\, dy - 2\hat{y}_{w^*} \int_{-\infty}^{\hat{y}_{w^*}} y f(y)\, dy + \int_{-\infty}^{\hat{y}_{w^*}} y^2 f_{Y|W=w^*}(y)\, dy\right) - a_l\left(\hat{y}_{w^*}^2 - 2\hat{y}_{w^*} \int_{-\infty}^{\infty} y f_{Y|W=w^*}(y)\, dy + \int_{-\infty}^{\infty} y^2 f_{Y|W=w^*}(y)\, dy\right) = 0. \tag{31}$$

Now, with any fixed $i = 1, 2, \ldots, m$, let the functions $\mathbf{U}_i$ and $\mathbf{V}_i$ defined by dependencies (16) and (25) be given, and furthermore $\mathbf{W}_i : \mathbf{R} \to \mathbf{R}$ be introduced as

$$\mathbf{W}_i(\hat{y}_{w^*}) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_{w^*}} y^2 \mathbf{K}_1\left(\frac{y - y_i}{h_1}\right) dy, \tag{32}$$

while the kernel $\mathbf{K}_1$ should be chosen so as to guarantee their useful analytic form. Making use of the above notations, condition (31) can be expressed in the following form:

$$\sum_{i=1}^{m} d_i[(a_l + a_r)(\hat{y}_{w^*}^2 \mathbf{U}_i(\hat{y}_{w^*}) - 2\hat{y}_{w^*} \mathbf{V}_i(\hat{y}_{w^*}) + \mathbf{W}_i(\hat{y}_{w^*})) + 2a_l \hat{y}_{w^*} y_i] - (a_l m \hat{y}_{w^*}^2 - \mathbf{W}_\infty) \sum_{i=1}^{m} d_i = 0, \tag{33}$$

where

$$\mathbf{W}_\infty = \int_{-\infty}^{\infty} y^2 \mathbf{K}_1(y)\, dy \tag{34}$$

is finite, on the basis of assumption (30). If the left-hand side of formula (33) is denoted as $L(\hat{y}_{w^*})$, then – also taking into account the equalities $\mathbf{V}_i'(\hat{y}_{w^*}) = \hat{y}_{w^*} \mathbf{U}_i'(\hat{y}_{w^*})$ and $\mathbf{W}_i'(\hat{y}_{w^*}) = \hat{y}_{w^*} \mathbf{V}_i'({}_{w^*})$ resulting from dependencies (16), (25), and (32) – the derivative of the function L is

$$L'(\hat{y}_{w^*}) = \sum_{i=1}^{m} d_i[2(a_l + a_r)(\hat{y}_{w^*} \mathbf{U}_i(\hat{y}_{w^*}) - \mathbf{V}_i(\hat{y}_{w^*})) + 2a_l y_i] - 2a_l m \hat{y}_{w^*} \sum_{i=1}^{m} d_i. \tag{35}$$

Finally, the desired estimator can be calculated numerically through Newton's algorithm (19)–(21), while the functions L and L' are given by dependencies (33)–(35).

The above investigations can be similarly transposed to a higher order of asymmetrical polynomial loss function (28), although on account of their extreme nature, they seem to be useful only for especially atypical applicational tasks.

4. Kernel used

In the linear case, presented in Section 3.1, the kernel $\mathbf{K}_1$ can be assumed in the Cauchy form

$$\mathbf{K}_1(x) = \frac{2}{\pi} \frac{1}{(1 + x^2)^2}. \tag{36}$$

Then

$$\mathbf{U}_i(\hat{y}_{w^*}) = \frac{\frac{\hat{y}_{w^*} - y_i}{h_1}}{\pi\left[1 + \left(\frac{\hat{y}_{w^*} - y_i}{h_1}\right)^2\right]} + \frac{1}{\pi} \operatorname{arctg}\left(\frac{\hat{y}_{w^*} - y_i}{h_1}\right) + \frac{1}{2}, \tag{37}$$

while if one applies the plug-in method recommended here, the constants there amount to $\int_{\mathbf{R}} x^2 \mathbf{K}_1(x)\, dx = 1$ and $\int_{\mathbf{R}} \mathbf{K}_1(x)^2\, dx = 5/4\pi$.

In the quadratic case (Section 3.2) also Cauchy kernel (36) is proposed; then formula (37) remains true and additionally

$$\mathbf{V}_i(\hat{y}_{w^*}) = y_i\left(\frac{\frac{\hat{y}_{w^*} - y_i}{h_1}}{\pi\left[1 + \left(\frac{\hat{y}_{w^*} - y_i}{h_1}\right)^2\right]} + \frac{1}{\pi} \operatorname{arctg}\left(\frac{\hat{y}_{w^*} - y_i}{h_1}\right) + \frac{1}{2}\right) - \frac{h_1}{\pi\left[1 + \left(\frac{\hat{y}_{w^*} - y_i}{h_1}\right)^2\right]}. \tag{38}$$

In the cubic case considered in Section 3.3, Cauchy kernel (36) must by modified here to the form

$$\mathbf{K}_1(x) = \frac{8}{3\pi} \frac{1}{(1 + x^2)^3}. \tag{39}$$

An increase of the power in the denominator has been implied with the necessity of ensuring the fulfillment of condition (30). Here

$$\mathbf{U}_i(\hat{y}_{w^*}) = \frac{\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^3}{\pi\left(1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right)^2} + \frac{5\,\frac{\hat{y}_{w^*}-y_i}{h_1}}{3\pi\left(1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right)^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right) + \frac{1}{2}, \tag{40}$$

$$\mathbf{V}_i(\hat{y}_{w^*}) = -\frac{2h_1}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + y_i\left(\frac{\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^3}{\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + \frac{5\,\frac{\hat{y}_{w^*}-y_i}{h_1}}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right) + \frac{1}{2}\right), \tag{41}$$

$$\begin{aligned}\mathbf{W}_i(\hat{y}_{w^*}) = {} & -\frac{4h_1y_i}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + y_i^2\left(\frac{\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^3}{\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + \frac{5\,\frac{\hat{y}_{w^*}-y_i}{h_1}}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right) + \frac{1}{2}\right) \\ & + h_1^2\left(\frac{\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^3}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} - \frac{\frac{\hat{y}_{w^*}-y_i}{h_1}}{3\pi\left[1+\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right)^2\right]^2} + \frac{1}{3\pi}\operatorname{arctg}\left(\frac{\hat{y}_{w^*}-y_i}{h_1}\right) + \frac{1}{6}\right).\end{aligned} \tag{42}$$

For kernel (39), the constants used within the plug-in method are $\int_{\mathbf{R}} x^2\mathbf{K}_1(x)\,dx = 1/3$ and $\int_{\mathbf{R}} \mathbf{K}_1(x)^2\,dx = 7/4\pi$.

The kernels for coordinates corresponding to the conditioning variable W can be assumed freely, as they are not used in any further analytical calculations. In particular they can be in the aforementioned Cauchy form, i.e. $\mathbf{K}_2 \equiv \mathbf{K}_3 \equiv \ldots \equiv \mathbf{K}_{n_W+1} \equiv \mathbf{K}_1$, where $\mathbf{K}_1$ is given by formula (36) or (39), respectively.

5. Experimental verification

The correct functioning and positive attributes of the algorithm presented in this paper were confirmed with detailed numerical and experimental verification.

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

Table 1
Values of estimator for the asymmetrical linear case (Section 3.1).

$\frac{a_l}{a_r}$		$\frac{1}{10}$	$\frac{1}{3}$	1	3	10
w^*	m					
0	10	−3.245 ± 0.641	−2.175 ± 0.574	−0.678 ± 0.818	1.339 ± 0.904	2.987 ± 0.853
	20	−3.226 ± 0.518	−2.144 ± 0.444	−0.734 ± 0.678	1.301 ± 0.710	2.852 ± 0.626
	50	−3.140 ± 0.342	−2.162 ± 0.297	−0.877 ± 0.490	1.254 ± 0.525	2.716 ± 0.439
	100	−3.082 ± 0.304	−2.131 ± 0.259	−0.805 ± 0.416	1.415 ± 0.389	2.689 ± 0.300
	200	−3.037 ± 0.217	−2.148 ± 0.158	−0.875 ± 0.348	1.454 ± 0.308	2.700 ± 0.239
	500	−2.991 ± 0.136	−2.139 ± 0.121	−0.906 ± 0.264	1.548 ± 0.201	2.700 ± 0.165
	1000	−2.974 ± 0.102	−2.136 ± 0.094	−0.919 ± 0.176	1.515 ± 0.167	2.678 ± 0.128
	∞	−2.955	−2.134	−0.937	1.545	2.657
1	10	−3.079 ± 0.663	−1.931 ± 0.703	−0.448 ± 0.947	1.258 ± 0.943	3.074 ± 0.798
	20	−3.070 ± 0.505	−2.002 ± 0.526	−0.576 ± 0.815	1.364 ± 0.817	2.939 ± 0.725
	50	−2.974 ± 0.395	−2.004 ± 0.350	−0.642 ± 0.574	1.446 ± 0.574	2.897 ± 0.463
	100	−2.942 ± 0.325	−2.020 ± 0.252	−0.699 ± 0.439	1.534 ± 0.409	2.848 ± 0.347
	200	−2.867 ± 0.213	−1.984 ± 0.177	−0.702 ± 0.321	1.623 ± 0.333	2.866 ± 0.255
	500	−2.841 ± 0.135	−1.984 ± 0.134	−0.781 ± 0.244	1.637 ± 0.226	2.827 ± 0.174
	1000	−2.827 ± 0.108	−1.993 ± 0.097	−0.790 ± 0.159	1.631 ± 0.182	2.793 ± 0.138
	∞	−2.798	−1.977	−0.804	1.684	2.809
2	10	−2.770 ± 0.912	−1.778 ± 1.040	−0.421 ± 1.096	1.604 ± 1.141	2.999 ± 0.891
	20	−2.875 ± 0.574	−1.822 ± 0.712	−0.382 ± 0.982	1.527 ± 1.029	3.044 ± 1.014
	50	−2.831 ± 0.449	−1.845 ± 0.405	−0.422 ± 0.700	1.529 ± 0.737	3.011 ± 0.609
	100	−2.775 ± 0.309	−1.856 ± 0.289	−0.471 ± 0.560	1.655 ± 0.550	3.014 ± 0.420
	200	−2.710 ± 0.206	−1.838 ± 0.185	−0.584 ± 0.353	1.740 ± 0.320	3.001 ± 0.265
	500	−2.686 ± 0.153	−1.839 ± 0.146	−0.657 ± 0.240	1.740 ± 0.272	2.952 ± 0.197
	1000	−2.679 ± 0.124	−1.835 ± 0.119	−0.654 ± 0.192	1.788 ± 0.197	2.959 ± 0.151
	∞	−2.643	−1.821	−0.670	1.825	2.962

Table 2
Values of estimator for the asymmetrical quadratic case (Subsection 3.2).

$\frac{a_l}{a_r}$		$\frac{1}{10}$	$\frac{1}{3}$	1	3	10
w^*	m					
0	10	−2.314 ± 0.709	−1.396 ± 0.707	−0.400 ± 0.769	0.691 ± 0.846	1.817 ± 0.925
	20	−2.346 ± 0.527	−1.431 ± 0.525	−0.455 ± 0.578	0.611 ± 0.612	1.708 ± 0.623
	50	−2.260 ± 0.353	−1.405 ± 0.350	−0.453 ± 0.394	0.612 ± 0.417	1.691 ± 0.405
	100	−2.194 ± 0.311	−1.338 ± 0.289	−0.381 ± 0.323	0.678 ± 0.344	1.714 ± 0.346
	200	−2.154 ± 0.221	−1.330 ± 0.230	−0.388 ± 0.260	0.663 ± 0.266	1.692 ± 0.245
	500	−2.141 ± 0.150	−1.320 ± 0.153	−0.381 ± 0.182	0.665 ± 0.193	1.663 ± 0.190
	1000	−2.129 ± 0.121	−1.326 ± 0.121	−0.394 ± 0.136	0.647 ± 0.138	1.641 ± 0.131
	∞	−2.113	−1.328	−0.400	0.641	1.622
1	10	−2.244 ± 0.759	−1.343 ± 0.798	−0.385 ± 0.896	0.666 ± 0.981	1.795 ± 1.055
	20	−2.118 ± 0.514	−1.219 ± 0.553	−0.245 ± 0.639	0.815 ± 0.699	1.909 ± 0.775
	50	−2.099 ± 0.357	−1.245 ± 0.329	−0.296 ± 0.382	0.756 ± 0.430	1.815 ± 0.458
	100	−2.014 ± 0.270	−1.179 ± 0.280	−0.235 ± 0.329	0.815 ± 0.363	1.841 ± 0.368
	200	−2.014 ± 0.231	−1.204 ± 0.230	−0.278 ± 0.259	0.762 ± 0.276	1.787 ± 0.274
	500	−2.010 ± 0.138	−1.210 ± 0.142	−0.276 ± 0.170	0.776 ± 0.179	1.790 ± 0.170
	1000	−1.990 ± 0.101	−1.185 ± 0.108	−0.250 ± 0.127	0.782 ± 0.131	1.791 ± 0.126
	∞	−1.963	−1.182	−0.265	0.772	1.749
2	10	−2.024 ± 0.938	−1.114 ± 1.001	−0.154 ± 1.074	0.869 ± 1.129	1.935 ± 1.169
	20	−1.902 ± 0.689	−1.015 ± 0.778	−0.073 ± 0.887	0.907 ± 0.815	1.987 ± 0.835
	50	−1.955 ± 0.489	−1.113 ± 0.507	−0.171 ± 0.581	0.873 ± 0.626	1.934 ± 0.628
	100	−1.908 ± 0.316	−1.073 ± 0.285	−0.128 ± 0.334	0.931 ± 0.385	1.971 ± 0.422
	200	−1.848 ± 0.236	−1.026 ± 0.238	−0.098 ± 0.281	0.962 ± 0.302	1.990 ± 0.301
	500	−1.838 ± 0.195	−1.030 ± 0.190	−0.102 ± 0.211	0.940 ± 0.212	1.947 ± 0.202
	1000	−1.833 ± 0.125	−1.046 ± 0.129	−0.123 ± 0.155	0.921 ± 0.167	1.930 ± 0.165
	∞	−1.816	−1.038	−0.128	0.907	1.887

$$E_1 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad Cov_1 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 4 \end{bmatrix}, \quad 50\%, \tag{43}$$

$$E_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad Cov_2 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \quad 20\%, \tag{44}$$

$$E_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad Cov_3 = \begin{bmatrix} l1 & 0.7 \\ 0.7 & 4 \end{bmatrix}, \quad 30\%. \tag{45}$$

In the case of factors (43) and (45), the describing variable Y and conditioning W are positively correlated. Factor (44), 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 and W as well as their standard deviations are

$$E_Y = -0.4, \quad \sqrt{V_Y} = \sigma_Y \cong 2.2, \tag{46}$$

$$E_W = 0, \quad \sqrt{V_W} = \sigma_W = 2. \tag{47}$$

The results acquired for asymmetrical linear (Section 3.1), asymmetrical quadratic (Section 3.2) and asymmetrical cubic (Section 3.3) cases, 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 ± standard deviation". The symbol ∞ denotes there 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 describing value σ_Y, i.e. 0.22 (see formula (46)), or where the standard deviation of the estimation error is greater than 20% of σ_Y, i.e. 0.44. One can note that the remaining (unshaded) results are for samples of sizes from 50 or 100. Taking into account that the distribution of the describing random variable X is trimodal, the need for such size seems reasonable in practice.

Analysis of particular columns of Tables 1–3 will now be discussed. In the symmetric case $a_l/a_r = 1$ presented – for comparison – in the middle column, the losses arising from under- and overestimation are the same.[1] The ratio $a_l/a_r = 1/3$ results in losses due to overestimation being 3 times greater than those from underestimation, which means that the estimator values decrease to reduce the probability of overestimation. For the case $a_l/a_r = 1/10$, where losses from overestimation are 10 times greater than those due to underestimation, this process is even more distinct, the values of the resulting estimator are even

[1] For the symmetrical linear case, the Bayes estimator is simply a median, and for the symmetrical quadratic – an expected value.

Table 3
Values of estimator for the asymmetrical cubic case (Section 3.3).

$\frac{a_l}{a_r}$ w^*	m	$\frac{1}{10}$	$\frac{1}{3}$	1	3	10
0	10	−1.924 ± 0.713	−1.055 ± 0.683	0.252 ± 0.704	0.573 ± 0.762	1.504 ± 0.867
	20	−1.915 ± 0.548	−1.090 ± 0.506	−0.319 ± 0.502	0.472 ± 0.520	1.353 ± 0.566
	50	−1.774 ± 0.380	−1.018 ± 0.349	−0.292 ± 0.343	0.456 ± 0.351	1.279 ± 0.373
	100	−1.691 ± 0.355	−0.954 ± 0.306	−0.245 ± 0.296	0.482 ± 0.306	1.268 ± 0.344
	200	−1.636 ± 0.230	−0.927 ± 0.214	−0.237 ± 0.212	0.475 ± 0.217	1.248 ± 0.239
	500	−1.629 ± 0.169	−0.927 ± 0.148	−0.248 ± 0.148	0.448 ± 0.157	1.193 ± 0.176
	1000	−1.610 ± 0.128	−0.924 ± 0.115	−0.255 ± 0.113	0.432 ± 0.116	1.165 ± 0.126
	∞	−1.559	−0.905	−0.257	0.411	1.120
1	10	−1.876 ± 0.781	−1.013 ± 0.764	−0.222 ± 0.796	0.596 ± 0.858	1.527 ± 0.960
	20	−1.683 ± 0.538	−0.869 ± 0.538	−0.103 ± 0.576	0.686 ± 0.641	1.570 ± 0.747
	50	−1.616 ± 0.407	−0.865 ± 0.358	−0.145 ± 0.353	0.597 ± 0.374	1.413 ± 0.416
	100	−1.506 ± 0.279	−0.787 ± 0.265	−0.091 ± 0.276	0.625 ± 0.297	1.400 ± 0.332
	200	−1.500 ± 0.243	−0.801 ± 0.226	−0.120 ± 0.227	0.585 ± 0.238	1.352 ± 0.267
	500	−1.489 ± 0.149	−0.798 ± 0.136	−0.121 ± 0.137	0.577 ± 0.142	1.330 ± 0.157
	1000	−1.465 ± 0.112	−0.778 ± 0.102	−0.107 ± 0.103	0.574 ± 0.110	1.322 ± 0.125
	∞	−1.414	−0.760	−0.112	0.555	1.259
2	10	−1.677 ± 0.901	−0.821 ± 0.910	−0.040 ± 0.948	0.763 ± 1.003	1.683 ± 1.093
	20	−1.485 ± 0.677	−0.680 ± 0.717	0.074 ± 0.779	0.806 ± 0.725	1.674 ± 0.810
	50	−1.461 ± 0.500	−0.723 ± 0.489	−0.010 ± 0.503	0.730 ± 0.527	1.549 ± 0.561
	100	−1.388 ± 0.341	−0.669 ± 0.308	0.031 ± 0.316	0.753 ± 0.349	1.538 ± 0.411
	200	−1.336 ± 0.256	−0.625 ± 0.237	0.066 ± 0.243	0.780 ± 0.263	1.556 ± 0.306
	500	−1.315 ± 0.208	−0.624 ± 0.187	0.050 ± 0.183	0.746 ± 0.188	1.497 ± 0.206
	1000	−1.316 ± 0.142	−0.633 ± 0.130	0.037 ± 0.132	0.728 ± 0.142	1.474 ± 0.165
	∞	−1.271	−0.615	0.033	0.699	1.408

smaller. The opposite effect takes place for $a_l/a_r = 3$, and to an even greater degree when $a_l/a_r = 10$ – the values of the estimator increase. It is worth adding that for extreme values of the ratio a_l/a_r, i.e. less than 1/10 and greater than 10, there may arise a natural necessity to enlarge the sample sizes – these cases, however, seem not to have any practical significance, as in such situations it is worth substituting the Bayes approach with the minimax [2].

Particular sections of rows, related to the subsequent conditioning values $w^* = 0$, $w^* = 1$, $w^* = 2$, will now be analyzed. The first of them corresponds to the modal value of the conditioning variable W (see formulas (43)–(45). Because the variables Y and W are correlated positively, for $w^* = 1$ the estimator values become bigger than for $w^* = 0$. When $w^* = 2$ these values increase even more. In the latter case the necessity of a larger sample size is evident, which seems to be justified as the conditioning value $w^* = 2$ lies on the second standard deviation of the variable W from its modal value. In the cases $w^* = -1$, $w^* = -2$ the estimator values are respectively smaller than for $w^* = 0$, which because of symmetry have been omitted from Tables 1–3. It should be underlined that the dependence of the estimator value on the given conditioning value w^*, considered in this paragraph, constitutes the essence of the conditional approach investigated in this paper.

The results in Tables 1–3 will now be collated, comparing the values obtained for the linear, quadratic and cubic cases, defined by loss functions (14), (22), and (28) with $k = 3$, respectively. Quadratic form (22) means that large estimation errors cause here significantly greater values of loss function than for the linear (14). The desired minimization of the expected value of losses requires the quadratic estimator to assume values which better "protect" against large errors, i.e. somewhat closer to an average value (median or expected value). The above becomes more distinct for the cubic case. A comparison of the results shown in Tables 1–3 confirms these suggestions for any ratio a_l/a_r and the conditioning value w^*. The property considered in this paragraph represents the quintessential possibilities arising from potential changes – in relation to a form obtained through fundamental research – the degree of polynomial in the preliminary form of the loss function (1).

Finally, it is worth noting that in any case shown in Tables 1–3, as the sample size increased, the obtained parameter value converged to the theoretical, and the standard deviation to zero. The above 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 calculational 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 proven 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. bounded supports, lack of data from the neighborhood of a given conditioning variable x^*, as well as the occurrence of discrete, binary and categorized coordinates of the conditioning variable W.[2]

[2] From the formal point of view this consists only of the introduction to the definition of kernel estimator (3) of additional factors related to coordinates of these types; for details see [19], [6]–Section 3.1.8 and/or [7]–Section 6.1.4), [20], respectively.

The concept presented in this paper was also verified experimentally by applying it to identification of a model of dynamic systems [21] submitted to robust control [22], based on the time-optimal control theory [23]. The idea of time-optimal control itself, ensuring a minimum of operating time, seems very attractive. In practice though this control, being extreme in nature (i.e. assuming maximal allowed values), proved to be highly sensitive to identification errors which were unavoidable in practice. It did however become a suitable basis for creating suboptimal structures, where such sensitivity would be eliminated. The task of parameter identification arising in this way remains in practice a problem of fundamental importance.

Consider a mechanical system with dynamics modeled by the differential inclusion

$$\ddot{y}(t) \in H(\dot{y}(t)) + u(t), \tag{48}$$

where y expresses the position of the object, u is a control with values limited to the interval $[u_*, u^*]$ and the function H, characterizing resistance to motion, is piecewise continuous and additionally multivalued at the points of discontinuity (particularly at zero it can represent phenomena connected with static friction). In the event of no resistance to motion, i.e. when $H \equiv 0$, inclusion (48) can be reduced to a differential equation $\ddot{y}(t) = m\,u(t)$ expressing the mass m submitted to the action of a force according to Newton's second law of dynamics. The above task constitutes therefore a problem of fundamental importance in the control of manipulators and robots. Object (48) was subjected to a robust control, which took on the values u_* or u^*, depending on where among the distinguished sets the system state was located; for details see the papers [24,25]. This concept was the basis for creating a complex algorithm for controlling a laboratory robot arm. Proper fixing of parameter values is of fundamental importance here, as in many cases underestimation results in sliding trajectories, with overestimation giving ineffective convergent cycles (or vice-versa), while both in the case of sliding trajectories as well as convergent cycles, their effectiveness depends on the assumed values of parameters. Some of these were conditioning in character: the parameters u_* and u^* (defining the maximum power of an actuator) depended on the interim power supply value and motion resistance value (described by the function H) to a large degree depended on current velocity. In particular cases the motion resistance value could additionally depend on position, and the mass m on the time passed since the last replenishment of substances consumed during the production cycle.

The use of the conditional approach of the Bayes parameter identification enabled a further reduction in operating time of about 5% with respect to the unconditional version, giving a time of about 10% less than that for results obtained using the mean value as an estimator, and about 5% shorter than in the case of the classical sliding mode control [26], where sliding trajectories had a notably larger number of switchings, which lowered their effectiveness.

Similar practical research was also carried out with success in medical applications, in establishing optimal dosages of anesthetic considering patients' body mass and general condition, as well as strategic sales in selecting policy for a mobile phone operator when negotiating with a business client characterized by many vastly different factors.

Generally it is worth stressing that in every case investigated, precision of the characteristics describing the parameter under investigation by providing the proper value for conditioning factors improved the result in proportion to the degree of differentiation of object features with respect to those factors. This occurred in the case of circumstantial changes in values for these factors, as well as structural object nonstationarity. In today's age of ever more available current measurement data and the possibility of instant inclusion in computer algorithms, this should be particularly underlined, providing real advantages from the procedure described in this paper.

6. Final remarks and summary

This paper presents the algorithm for calculating the conditional value of a parameter, ensuring a minimum expected value of losses with their asymmetrical form representing different results of over- and underestimation. The conditional approach allows in practice for refinement of the model by including the current value of the 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 numerical simulation as well as practical problems from control engineering, medicine and marketing.

An important question is the calculational 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, and the algorithm for its modification – recommended in Section 2 – have for every $n_W + 1$ coordinates of the variable X calculational complexity $O(m^2)$, but they are carried out once at the beginning of the procedure. The second phase consists of finding solutions of criterions (15), (23), (31) or more generally (29), with the aid of Newton's algorithm (19)–(21), 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 in type. Newton's algorithm in investigated cases most often required 6–10 iterations. This implies a relatively short calculation time for the second phase which means, 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.

Finally, it is worth adding that the concept developed here can be generalized to a multidimensional case, i.e. where the vector of conditionally correlated parameters is identified. However, in this 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; (for the unconditional case see [27]). Similarly it is possible to assume loss function (1) in an asymmetrical form of different degree of polynomial for negative and positive estimation errors. However such a case seems to have only theoretical significance, with no applicational connotations.

Acknowledgments

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 [28].

References

[1] E. Walter, I. Pronzato, Identification of Parametric Models, Springer-Verlag, Berlin, 1997.
[2] E.L. Lehmann, Theory of Point Estimation, Wiley, New York, 1983.
[3] S. Brandt, Statistical and Computational Methods in Data Analysis, Springer-Verlag, New York, 1999.
[4] P. Kulczycki, O. Hryniewicz, J. Kacprzyk (Eds.), Techniki informacyjne w badaniach systemowych, WNT, Warsaw, 2007.
[5] J.O. Berger, Statistical Decision Theory, Springer-Verlag, New York, 1980.
[6] P. Kulczycki, Estymatory jadrowe w analizie systemowej, WNT, Warsaw, 2005.
[7] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[8] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman and Hall, London, 1995.
[9] P. Kulczycki, M. Charytanowicz, Warunkowa bayesowska identyfikacja parametryczna z niesymetrycznymi stratami, *XVII Krajowa Konferencja Automatyki*, Kielce (Poland), 19–22 June 2011, pp. 215–226, post-conference publication – in press.
[10] A.J. Izenman, Recent developments in nonparametric density estimation, J. Am. Stat. Assoc. 86 (1991) 205–224.
[11] P. Kulczycki, Statistical inference for fault detection: a complete algorithm based on kernel estimators, Kybernetika 38 (2002) 141–168.
[12] P. Kulczycki, Estymatory jadrowe w badaniach systemowych, in: P. Kulczycki, O. Hryniewicz, J. Kacprzyk (Eds.), Techniki informacyjne w badaniach systemowych, WNT, Warsaw, 2007, pp. 79–105.
[13] P. Kulczycki, Kernel estimators in industrial applications, in: B. Prasad (Ed.), Soft Computing Applications in Industry, Springer-Verlag, Berlin, 2008, pp. 69–91.
[14] P. Kulczycki, M. Charytanowicz, P.A. Kowalski, S. Lukasik, The complete gradient clustering algorithm: properties in practical applications, J. Appl. Statist. 39 (2012) 1211–1224.
[15] P. Kulczycki, K. Daniel, Metoda wspomagania strategii marketingowej operatora telefonii komorkowej, Przeglad Statystyczny 56 (2009) 116–134.
[16] P. Kulczycki, P.A. Kowalski, Bayes classification of imprecise information of interval type, Control Cybern. 40 (2011) 101–123.
[17] P. Kulczycki, J. Waglowski, On the application of statistical kernel estimators for the demand-based design of a wireless data transmission system, Control Cybern. 34 (2005) 1149–1167.
[18] D. Kincaid, W. Cheney, Numerical Analysis, Brooks/Cole, Pacific Grove, 2002.
[19] I.A. Ahmad, P.B. Cerrito, Nonparametric estimation of joint discrete-continuous probability densities with applications, J. Statist. Plan. Inf. 41 (1994) 349–364.
[20] D.M. Titterington, A.W. Bowman, A comparative study of smoothing procedures for ordered categorical data, J. Stat. Comput. Simul. 21 (1985) 291–312.
[21] T.S. Soderstrom, P. Stoica, System Identification, Prentice Hall, Englewood Cliffs, 1994.
[22] U. Mackenroth, Robust Control Systems, Theory and Case Studies, Springer, Berlin, 2004.
[23] E.G. Sontag, Mathematical Control Theory, Springer, New York, 1998.
[24] P. Kulczycki, Fuzzy controller for mechanical systems, IEEE Trans. Fuzzy Syst. 8 (2000) 645–652.
[25] P. Kulczycki, R. Wisniewski, Fuzzy controller for a system with uncertain load, Fuzzy Sets Syst. 131 (2002) 185–195.
[26] V.I. Utkin, Sliding Modes in Control Optimization, Springer-Verlag, Berlin, 1992.
[27] A. Mazgaj, Modelowanie niepewnosci parametrow obiektu dla potrzeb sterowania optymalnego, Ph.D.-thesis (supervisor P. Kulczycki), AGH University of Science and Technology, Cracow, 2005.
[28] P. Kulczycki, A. Mazgaj, Parameter identification for asymmetrical polynomial loss function, Information Technology and Control, vol. 38, pp. 51–60; Errata: vol. 38, 2009, pp. 167–168.